2019

Machine Learning for Load Profile Data Analytics and Short-term Load Forecasting

Md. Rashedul Haq
South Dakota State University

MACHINE LEARNING FOR LOAD PROFILE DATA ANALYTICS AND SHORT-TERM LOAD FORECASTING

BY

MD. RASHEDUL HAQ

A thesis submitted in partial fulfillment of the requirements for the

Master of Science

Major in Electrical Engineering

South Dakota State University

2019

MACHINE LEARNING FOR LOAD PROFILE DATA ANALYTICS AND SHORT-TERM LOAD FORECASTING

MD. RASHEDUL HAQ

This thesis is approved as a creditable and independent investigation by a candidate for the Master of Science in Electrical Engineering degree and is acceptable for meeting the thesis requirements for this degree. Acceptance of this thesis does not imply that the conclusions reached by the candidates are necessarily the conclusions of the major department.

Zhen Ni, Ph.D.
Thesis Advisor — Date

George Hamer, Ph.D.
Head, Electrical Engineering and Computer Science — Date

Dean, Graduate School — Date

ACKNOWLEDGEMENTS

Firstly, I would like to express my sincere gratitude to the committee chair and my thesis advisor Dr. Zhen Ni for his continuous support during my Master's study. Its my immense pleasure to express deepest appreciation for his persuasive advice and interactive supervision. Without his enlightened supervision and continuous encouragement this thesis would not have been accomplished.

Beside advisor, the author of this thesis would like to thank the other committee members: Dr. Kwanghee Won, Dr. Marco Ciarcia, and Dr.Joy Scaria , for their insightful comments and precious advice. The author would also like to thank the Electrical Engineering and Computer Science Department of South Dakota State University for the financial support granted through Master's study.

The author would like to express deep hearted appreciation to the fellow research group members for their stimulating insightful discussions, for the sleepless nights we were working together before deadlines, and for all the fun we have had in the last two years.

The author is grateful to his beloved parents and siblings, who has provided moral and emotional support in the whole life. In particular, I am also grateful to Dr. Md. Shahjahan for enlightening me the first glance of research.

CONTENTS

ABBREVIATIONS

AEMO	Australian Energy Market Operator
ANN	Artificial neural network
BPNN	Back propagation neural network
DBN	Deep belief network
ELM	Extreme learning machine
EMD	Empirical mode decomposition
ENN	Extended Nearest Neighbor
IEMD	Improved empirical mode decomposition
IMF	Intrinsic mode function
kNN	k Nearest Neighbor
MAE	Mean absolute error
MAPE	Mean absolute percentage error
MLP	Multi layer perceptron
NEM	National Energy Market
NSW	New South Wales
RBFNN	Radial basis functions neural networks
RBM	Restricted Boltzmann Machine
RMSE	Root mean square error
SG	Smart Grid
SME	Small medium enterprise

SOM	Self organizing map
STLF	Short-term Load Forecasting
SVM	Support vector machine
VaR	Value at Risk

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

MACHINE LEARNING FOR LOAD PROFILE DATA ANALYTICS AND SHORT-TERM LOAD FORECASTING

MD RASHEDUL HAQ

2019

Short-term load forecasting (STLF) is a key issue for the operation and dispatch of day ahead energy market. It is a prerequisite for the economic operation of power systems and the basis of dispatching and making startup-shutdown plans, which plays a key role in the automatic control of power systems. Accurate power load forecasting not only help users choose a more appropriate electricity consumption scheme and reduces a lot of electric cost expenditure but also is conducive to optimizing the resources of power systems. This advantage helps while improving equipment utilization for reducing the production cost and improving the economic benefit, and improving power supply capability. Therefore, ultimately achieving the aim of efficient demand response program. This thesis outlines some machine learning based data driven models for STLF in smart grid. It also presents different policies and current statuses as well as future research direction for developing new STLF models. This thesis outlines three projects for load profile data analytics and machine learning based STLF models.

First project is, load profile classification and determining load demand variability with the aim to estimate the load demand of a customer. In this project load profile data collected from smart meter are classified using recently developed extended nearest neighbor (ENN) algorithm. Here we have calculated generalized class wise statistics which will give the

idea of load demand variability of a customer. Finally the load demand of a particular customer is estimated based on generalized class wise statistics, maximum load demand and minimum load demand.

In the second project, a composite ENN model is proposed for STLF. The ENN model is proposed to improve the performance of k-nearest neighbor (kNN) algorithm based STLF models. In this project we have developed three individual models to process weather data i.e., temperature, social variables, and load demand data. The load demand is predicted separately for different input variables. Finally the load demand is forecasted from the weighted average of three models. The weights are determined based on the change in generalized class wise statistics. This projects provides a significant improvement in the performance of load forecasting accuracy compared to kNN based models.

In the third project, an advanced data driven model is developed. Here, we have proposed a novel hybrid load forecasting model based on novel signal decomposition and correlation analysis. The hybrid model consists of improved empirical mode decomposition, T-Copula based correlation analysis. Finally we have employed deep belief network for making load demand forecasting. The results are compared with previous studies and it is evident that there is a significant improvement in mean absolute percentage error (MAPE) and root mean square error (RMSE).

CHAPTER 1 INTRODUCTION

Nowadays, daily operations and planning in a smart grid require a day-ahead load forecasting of customers. The accuracy of day-ahead load-forecasting models has a significant impact on many decisions such as scheduling of fuel purchases, system security assessment, economic scheduling of generation capacity, and planning for energy transactions. However, day-ahead load forecasting is a challenging task due to its dependence on external factors such as meteorological and exogenous variables. To this end, it is important to reduce uncertainty associated with demand, and it is important that load demand forecast is as accurate as possible. To achieve this, it is necessary to know the features of the load demand to be forecasted and, based on this, the purpose of this thesis is to develop or to choose the best and the most accurate model for short-term load forecasting (STLF).

1.1 Background

Energy management systems are designed to monitor, optimize, and control the smart grid energy market. Demand-side management, considered as an essential part of the energy management system, can enable utility market operators to make better management decisions for energy trading between consumers and the operator [1], [2]. In this system, a priori knowledge about the energy load pattern (e.g., day-ahead forecasted load) can help reshape the load and cut the energy demand curve, thus allowing a better management and distribution of the energy in smart grid energy systems. Designing a computationally intelligent load forecasting model is often a primary goal of energy management system. The accurate electricity load forecasting has a significant role in power system. It is useful for making optimal decisions to ensure the secure, reliable and economic operation of the power system [3]–[5]. Depending on the forecast horizon load forecasting can be classified into short-term, medium-term, and long-term [6]. Usually STLF predicts the hourly (or half hourly) load demand for next few hours to next few days. The results of STLF are used for short-term operational planning of the power system, e.g., generation scheduling. Day-ahead load forecasting falls under STLF class, which aims to predict the next-day's load demand of each dispatching interval. When load forecasting is intended for longer time interval i.e., from next few months to next few years, then medium-term and long-term load forecasting models are developed. The results of medium-term and long term load forecasting are linked to system mid-term and long-term planning practices, e.g., component maintenance scheduling or generation/transmission expansion planning. The applications of different types of load forecasting models are illustrated as below:

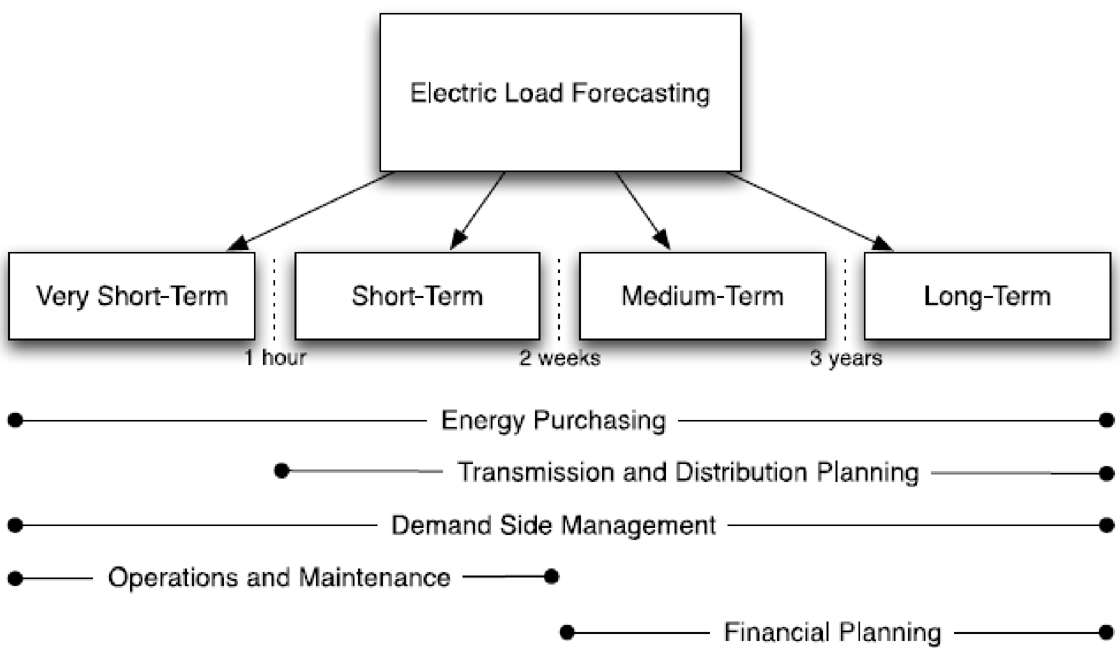

Figure 1.1. Electric load forecasting applications and classification [6].

In the modern world, with the high development of electricity market and rapid expansion of power system, short-term load forecasting is becoming an important factor of power system operation scheduling. In order to increase the load forecasting accuracy, many researchers have proposed different data driven methods. Smart meter is one of the most important devices implemented in the smart grid (SG) . With smart meters, electrical data such as voltage and frequency are measured and real-time energy consumption information is recorded. Smart meter supports bidirectional communications between the meter and the central system. Also, the smart meter has the built in ability to disconnect and reconnect certain loads remotely, which can be used to monitor and control the user's devices and appliances so as to manage demands and loads within the "smart-buildings" in the future. Thus, how to improve the accuracy of short-term load forecasting has always been the focus of load forecasting study for this thesis.

1.2 Related Work on Machine Learning Based Data Driven STLF Models

Load demand forecasting is one of the dominant policy tools used by decision-makers in the energy sector. Forecasts are essential for planning, constructing strategies, setting policies, and risk management, and constitute one of the main factors in pricing. Therefore, it is imperative to model load accurately. The financial costs of forecast errors are so high that much research is focused on reducing the error even by a fraction of a percentage point [7]. Electric power utilities need accurate load forecast model for reliable and efficient planning and operation of the grid. For short-term load forecasts such as day-ahead case, data-driven mathematical models have been traditionally studied with various machine learning methods [8].

The problem of STLF can be viewed as a time-series prediction problem (e.g.,[9]) in which the load is predicted based on the current-day load. Moreover, electricity load demand depends on several factors including weather, time, and socio-economic constraints [10]. Not considering the date, temperature, and other weather influences, such models can produce poor forecasting performance. When heterogeneous exogenous variables are considered as input for the load forecasting model, the STLF problem becomes complex. The STLF problem has been addressed with different methods in the literature such as regression [11], fuzzy logic [12], artificial neural network (ANN) [13]–[15] and exponential smoothing methods [16]. Each has many different models, for example, the regression based method includes auto-regressive moving average [17], autoregressive integrated moving average [18], and support vector regression [19]. Likewise, ANN based method includes bagged ANN, cascaded ANN [20], radial basis functions neural networks (RBFNN) [21], back propagation neural network (BPNN) [22] and extreme learning machine (ELM) [23]. ANN based models are the most popular methods. Rui Zhang et.al., [23] combined a series of ANN models for developing the ELM to further improve the load forecasting accuracy. Neural network based models have been widely used in load forecasting models due to their learning capability of complex nonlinear relationships between load demand and the effects of historical data [24]. However, neural network based models are associated with some potential drawbacks: overfitting of the model, sensitivity to random weight initialization, and tendency to converge toward local optima [25]. Later, in 2016 Rui Zhang et.al., [26] proposed a composite k-nearest neighbor (kNN) model for day-ahead load forecasting with temperature forecasts. Authors have provided a computationally efficient & simple model compared to their previous work [23]. But it is noteworthy that neighbor selection based on kNN is affected by irrelevant features which affects the load forecasting accuracy. Therefore, there is still some room to further increase the load forecasting accuracy.

1.3 Related Work on Advanced Data Driven Hybrid STLF Models

Over the last few years, researchers have proposed many models to forecast electricity load for varying time interval. Depending on the forecast time interval, those models can be classified into short-term, medium-term, and long-term load forecasting model [6]. Usually short-term load forecasting (STLF) models predict the half hourly or hourly load demand for next few hours to next few days. The results of STLF are used for short-term operational planning of the power system, e.g., generation scheduling. When load forecasting is intended for longer time interval i.e., from next few months to next few years, then medium-term and long-term load forecasting models are developed. The results of medium-term and long term load forecasting are linked to system mid-term and long-term planning practices, e.g., component maintenance scheduling or generation/transmission expansion planning. Based on the model architecture, load forecasting models are primarily divided into two classes: traditional statistical models and advanced data driven models. Traditional statistical models are built using linear regression function where the problem of STLF is viewed as a time-series prediction problem [9], [11]. The regression based models include auto-regressive moving average [17], autoregressive integrated moving average [18], autoregressive moving average with exogenous variable [27] and support vector regression [19]. The regression based models are effective for predicting stationary time series. However, load demand time series is non-stationary and shows nonlinear characteristics, thus advanced data driven models are proposed in recent times. To date, STLF problem has been investigated with different advanced data driven models. Advanced data driven models include: fuzzy logic based [12], artificial neural network (ANN) based [13], [14] and exponential smoothing methods [16]. ANN based models are the most popular among advanced data driven models. The ANN based method includes: bagged ANN, cascaded ANN [20], radical basis functions neural networks [21], back propagation neural network [22] and extreme learning machine [23]. Both statistical and advanced data driven models are proposed to predict the load demand. However, a single model is inadequate to represent inherent characteristics of electricity load demand because it depends on several factors including weather, time, and socio-economic constraints [10]. If we do not consider the date, temperature, and other weather influences, such models produce fair forecasting performance. When heterogeneous external factors are considered as input for the load forecasting model, the STLF problem becomes complex.

Thus, hybrid models are formed by integrating different models for improving the forecasting accuracy. The reason is that, different models can capture the features of electricity load profiles. In general, the hybrid models are classified into two main categories. For the first category model, electricity load is predicted separately by different models [28]–[33]. For the second category model, electricity load is decomposed into several components. Then each component is predicted by a suitable model [34]–[40]. For both of these category model, we still need to look for further advanced data driven models.

1.4 Motivations and Contributions

Electricity load forecasting is essential for the utility provider to manage the demand response program efficiently in day ahead energy market. From the information of electricity load demand of consumers, utility providers can estimate how much electric energy is needed in the grid. The objective of the utility provider is to minimize the cost of energy production and purchasing [41]. In this scenario, a prior knowledge about the energy demand can help utility providers to make proper planning of generation units schedule and amount of energy to be purchased [5]. Cost of energy production and purchasing is reduced in following way: (i) proper generation scheduling will save loss from running of extra generating unit, and (ii) electricity purchasing cost is minimized because power plants sell electrical energy at lower cost if bought in advance. The accurate electricity load forecasting has a significant role in power system, but any error in forecast incurs additional cost. According to Bunn and Farmer [42], [43] an increase of forecasting error of 1% caused an increase of $13 million in operating costs per year for one electric utility operator in the United Kingdom. Power grid planning, investment and transaction are also based on accurate electricity load forecasting. Thus, accurate electricity load forecasting is prerequisite for making secure, reliable and economic operation of power system [3].

Motivated from the work in [44] and to tackle the above mentioned limitations in traditional classification approaches, the first project investigates recently developed classification algorithm called extended nearest neighbor (ENN) [45] to classify customers electricity load profile data. ENN algorithm is different from traditional classification approaches because it separates different load profiles from maximum gain of intra-class coherence. The maximum gain of intra-class coherence is learned from global distribution of all available training samples. In this classification approach we have computed the generalized class wise statistics to predict the load demand variability of a particular customer. We have also provided an insight to estimate the load demand of particular customer. Thus we have adopted a load profile classification approach at more granular level to facilitate load monitoring control units.

Motivating from works in [23], [25], [26], we propose a novel load forecasting model in the second project. To solve the STLF problem and further improve the forecasting accuracy, the proposed solution in this project first classifies the load profile data into different classes according to various seasons of the year. The ENN is incorporated to solve the problem of kNN based models. The proposed composite model in this project consists of three different individual models and the result of the models are combined together by tuned weight factor for making a final forecasting output. The contribution of this project is to suppress the influence of irrelevant features and determining the weights for individual models without incurring additional computational complexity.

Motivating from the works in [28]–[40], we have proposed a novel hybrid load forecasting model which includes new signal decomposition technique and new correlation analysis technique for the third project. To mitigate end effect and envelope fitting limitation associated with traditional empirical mode decomposition (EMD), a new improved empirical mode decomposition (IEMD) method is proposed. By using IEMD, the original load demand time series is decomposed into several low frequency components to extract the characteristics of electricity load more accurately and effectively. Later on, to compensate for the information loss during signal decomposition, the effect of exogenous external factors (i.e., weather variables) is incorporated in the forecasting model. To accomplish this task, we have introduced new correlation analysis technique i.e., T-Copula for: (i) determining the interdependence between electricity load and exogenous external factors, and (ii) deriving the peak load indicative threshold parameters from value at risk (VaR).

1.5 The Structure of Thesis

The rest of the thesis is organized as follows. In Chapter 2, load profile classification using ENN is discussed. In this chapter we have also predicted the load demand variability of individual customers with an insight to estimate the load demand of a customer. In the Chapter 3, we have presented the composite ENN model for STLF which consists of three models. In the Chapter 4, we have presented the novel hybrid STLF model. The hybrid model consists of improved empirical decomposition, correlation analysis and deep belief network. In this work, we have solved the end effect and envelope fitting limitation of traditional empirical mode decomposition based signal decomposition. In this work, the accuracy of load forecasting is improved significantly. Finally, conclusions of the thesis and possible future works are presented in Chapter 5.

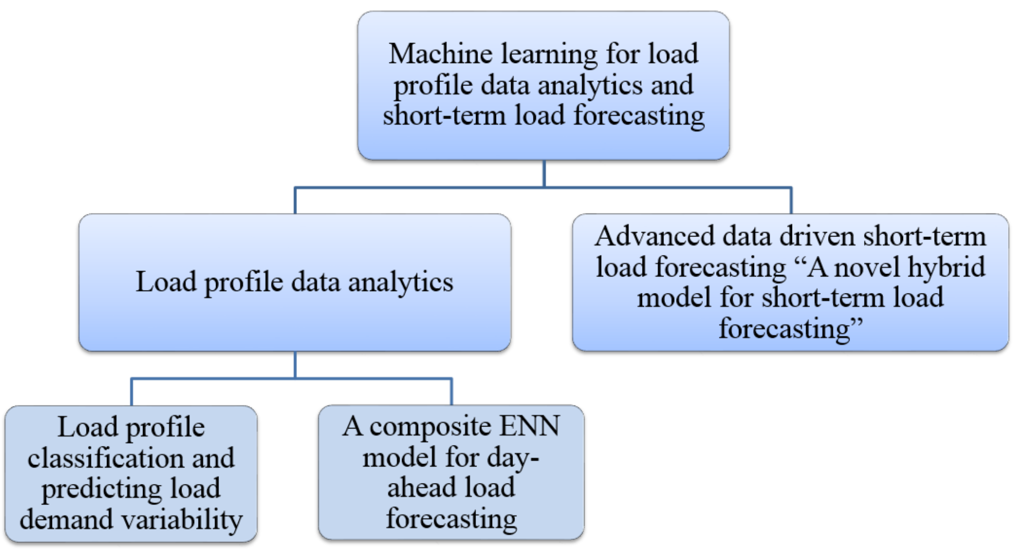

Figure 1.2. The structure of thesis.

CHAPTER 2 Load Profile Classification and Predicting Their Variability

2.1 Overview

Large volume of electricity load profile data collected from smart meter reveal information about customer's electricity consumption. The precise knowledge of customer's load profile classification help the load service entities for the efficient management of demand response program, predicting load demand variability, estimating load demand, and energy efficiency improvement. In this chapter, a recently developed classification approach called extended nearest neighbor (ENN) is introduced to classify customer's electricity load profile data. This classification approach predicts the load demand variability of customer and then estimate the load demand of a customer. Instead of focusing on the shape of load curves, this classification approach identifies customer class from the maximum gain of intra-class coherence. By exploiting intra-class coherence from generalized class-wise statistic of all available training samples, the ENN algorithm learns from global distribution to improve classification accuracy. The load variability of a customer is predicted based on the generalized class-wise statistics. Finally the future load demand of a particular customer is estimated from customer previous load profile history. The proposed method is validated on 21600 smart metered customers electricity load profile data collected from United States of America. Numerical results from case study verifies that, ENN algorithm provides higher classification accuracy than other comparative algorithms. The performance of load demand estimation is evaluated based on mean absolute percentage error (MAPE). The MAPE value for the case study is found to be 4.98%, which indicates higher estimation accuracy.

2.2 Introduction

Load profile classification plays a vital role in the time of use tariff design [46], predicting the load demand variability of a customer, estimating the load demand or customer scale load forecasting [47], demand response & energy efficiency targeting [48], and non-technical loss detection [49]. Better understanding of electricity consumption variability from load profile classification can help utility operators to adjust customer's strategies more economically and optimally to participate in demand response programs. Heretofore, in literature various methods have been proposed for classifying electricity load profile data of customers. Those methods include k-means [50], fuzzy k-means [51], hierarchical clustering [52], [53], self-organizing maps (SOM) [54] , support vector machines [55] , subspace clustering [56] etc. Traditional load profile classification techniques are classified into direct and indirect classification. In direct classification methods, load profile data is applied directly for classification without any data dimension reduction. But for indirect classification, preprocessed data i.e., sparse signal consisting of extracted features is applied for classification [44], [57]–[59]. Directly applying previous classification methods to smart meter data has limitation of either increased computational burden or fair accuracy toward high dimensional input feature space. An essential prerequisite for traditional load profile classification is the sufficient similarity among load profile curves. In this case, traditional load profile classification approaches considers the distance between load profiles for similarity measure. Since load profile of individual customers are volatile thus different factors e.g., sudden spike or tiny time shift, magnitudes will interfere with each other during classification. Besides, deluge of electricity consumption data collected at larger frequency introduces challenges of computational burden and accuracy. Antti Mutanen et.al., proposed customer classification and load profiling method for distribution systems [44]. In their work, load profile classification method includes k-nearest neighbor (kNN), which classify the samples based on distance measurement between load profiles. Even though Antti Mutanen et.al., has reported promising classification results but classifying customers electricity load profile data directly from mean distance measurement results in fair classification accuracy. The fair classification accuracy is due to sensitiveness of distance measurements. From the work proposed in [44], it is not possible to predict load demand variability of a customer and authors did not documented any insight to estimate the load demand of a customer.

Motivated from the work in [44] and to tackle the above mentioned limitations in traditional classification approaches, our objective is to investigates recently developed classification algorithm called extended nearest neighbor (ENN) [45] to classify customers electricity load profile data. ENN algorithm is different from traditional classification approaches because it separates different load profiles from maximum gain of intra-class coherence. The maximum gain of intra-class coherence is learned from global distribution of all available training samples. In this classification approach we have computed the generalized class wise statistics to predict the load demand variability of a particular customer. We have also provided an insight to estimate the load demand of particular customer. Thus we have adopted a load profile classification approach at more granular level to improve classification accuracy. This work is different from our recent work [60], where we have proposed composite extended nearest neighbor model for day ahead load forecasting. In that work, we have developed three different models and final load is forecasted from weighted sum of three model's output.

2.3 Existing Load Profile Classification Approaches and Feature Extraction

Heretofore, many researchers have reported different classification methods in literature for classification of electricity load profile data. Among those methods, neural network, kNN, support vector machine (SVM) based electricity load profile classification are common. In this chapter the performance of load profile classification using ENN is compared with those methods. For review, brief introduction of comparative study algorithms are presented below:

2.3.1 Neural Network

Neural network based classification is a supervised learning. It is an iterative learning process in which inputs are presented to the network one at a time. During the learning phase, the network is trained by adjusting the weights to predict the correct class label of incoming sample. Neural network is commonly used as a general mapping between the input features and output variables. In this classification approach, a multi-layer perceptron is adopted with hidden layer to build relationship between the features and corresponding labels. Advantages of neural network is the high tolerance toward noisy data as well as their ability to classify patterns on which they have not been trained. In 2015, Jamie Buitrago et. al., have applied neural network for classification of electricity load profile [61]. Neural network classification yields fair accuracy and more computational burden for massive, volatile and uncertain electricity load profile data. Computational time is also increased for high dimensional load profile feature space. Let input x_i (i is the index of the sample) includes n number of features. The output o_i corresponding to input sample x_i using neural network is follows [62],

$$o_i = \phi(\omega . x_i) \tag{2.1}$$

where ϕ is the sigmoid function and ω is the weights of neural network. The prediction error is given by,

$$\hat{e} = \sum_{i=1}^{n} \frac{1}{2}(t_i - o_i)^2 \tag{2.2}$$

where t_i is the true class label of the $i-$th sample. Weight is updated using gradient descent method in following way,

$$\Delta\omega = -\eta \frac{\partial \hat{e}}{\partial o} \frac{\partial o}{\partial \phi} \frac{\partial \phi}{\partial \omega} \tag{2.3}$$

here η represents learning rate. Here our task is to predict the class of a sample i.e., $f(o_i)$. It can be predicted from the value of o_i after using winner take all rule.

2.3.2 k-Nearest Neighbor

kNN is an instance based supervised learning used for classification and regression. The input consists of k nearest neighbor i.e., user defined parameter k. An object is classified by the majority vote of its nearest neighbors. Nearest neighbors are determined based on distance measurements. There are many distance measurement methods: Euclidean, Manhattan and Minkowski distances. Recently, Antti Mutanen et.al., and Asadi Majd et. al., have applied kNN for load profile classification [44], [63]. kNN is easy to implement and has less computational burden but classification accuracy is influenced by irrelevant features. This means if majority of the nearest neighbors of a test sample comes from another class (i.e., different from true class of test sample), then the test sample will be misclassified.

2.3.3 Support Vector Machine

SVM is a supervised classification approach that performs classification tasks by constructing a hyperplane in multidimensional feature space to separate the samples of different class labels. Classification using SVM is based on finding the hyperplane that gives the maximum margin for separating training examples. For finding best separating hyperplane SVM searches for closest points i.e., support vectors. The linear SVM is formulated as an optimization problem as follows [64],

$$\begin{aligned} &\min_{\gamma,\omega,b} \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{m} \xi_i \\ &s.t. y_i(\omega^T x_i + b) \geq 1 - \xi_i \end{aligned} \tag{2.4}$$

where x_i represents features, ω denotes the weight of features; $C > 0$ is penalty parameter for training error; ξ_i denotes loss function; b is the bias term in SVM. When the optimal value of C and weight feature ω have been found, the prediction of class label of a sample is done as follows,

$$f(a_i) = \text{sgn}(\omega^T.x_i + b) \tag{2.5}$$

Nowadays SVM classifier is intriguing among researchers. Vignesh V et. al., classified load consumption data using support vector machine [65]. For high dimensional input feature space classification using SVM results increased computational burden.

2.3.4 Feature extraction

Since the computational burden of SVM classification approach toward high dimensional input data is increased, researchers use data dimension reduction techniques [44], [57], [58]. In most of current work, researchers are being using principle component analysis for feature extraction [44]. Principle component analysis is a statistical procedure that uses orthogonal transformation to convert a set of observations of possibly correlated variables into a set of linearly uncorrelated variables called principle components. Additionally in some other work, peak load demand and load adjustment is also used as features in sparse representation of a signal [58].

2.4 Extended Nearest Neighbor Algorithm

2.4.1 What is ENN

ENN is a new supervised classification method that makes classification based on the maximum gain of intra-class coherence [45]. ENN makes a prediction in a two-way communication style: it considers not only the nearest neighbor of test sample but also who (training sample) considers the test sample as their nearest neighbor. ENN is a nonparametric classification approach which is more suitable toward high dimensional input data. ENN algorithm is different from kNN algorithm because ENN algorithm doesn't consider only the nearest neighbor for class prediction. ENN algorithm makes prediction from the maximum gain of intra class coherence.

2.4.2 Mathematical Formulation

In the proposed ENN classification approach, the first step is to determine generalized class-wise statistic. It is calculated from global distribution of all available training sample which gives the idea about the distribution of dataset i.e., whether they are mixed well or widely separated. A large value of generalized class-wise statistic means that samples for that particular class are more closer together whereas smaller value indicates opposite. A class with larger generalized class-wise statistic means that nearest neighbors for that class are dominated by same class samples. Note that generalized class-wise statistics lies in the range of 0 to 1. For each class of data samples the generalized class-wise statistic T_i for class i is determined as [45],

$$T_i = \frac{1}{n_i k} \sum_{x \in S_i} \sum_{r=1}^{k} I_r(x, S = S_1 \cup S_2 \ldots S_i) \tag{2.6}$$

where i represents the label of class 1,2,...etc., S_1, S_2,...S_i denotes corresponding set of data samples in each of the class, x denotes one single sample in class S_i and k is user defined parameter which represents the number of nearest neighbor. In this equation indicator function $I_r(x,S)$ is used to check whether both the sample x and its $r-th$ nearest neighbor belongs to the same class or not. Indicator function $I_r(x,S)$ is defined as,

$$I_r(x,S) = \begin{cases} 1, & \text{if} x \in S_i \text{and} NN_r(x,S) \in S_i \\ 0, & \text{otherwise} \end{cases} \tag{2.7}$$

where $NN_r(x,S)$ represents $r-th$ nearest neighbor of sample x in S. The outcome of indicator function is 1 if both the sample x and it's $r-th$ nearest neighbor belongs to the same class. Conversely for dissimilarity between the sample and it's nearest neighbor class, indicator function outcome is zero. The nearest neighbors are determined based on Euclidean distance measurement. The Euclidean distance between each testing sample to all training samples are determined. These distances are arranged in ascending order to determine

1st, 2nd, 3rd, etc., nearest neighbor. Other distance measurement methods eg., Manhattan, Minkowski can also be used.

Let an unknown test sample to be classified i.e., testing phase. In testing phase unknown test sample is iteratively assigned to each possible classes. During each iteration, the generalized class-wise statistic is updated. For updating new generalized class-wise statistic T_i^j, following equation is considered,

$$T_i^j = \frac{1}{n_i' k} \sum_{x \in S_{i,j}'} \sum_{r=1}^{k} I_r(x, S' = S_1 \cup S_2 \ldots S_i \cup Z) \tag{2.8}$$

where n_i' is the size of $S_{i,j}'$, and $S_{i,j}'$ is defined as follows,

$$S_{i,j}' = \begin{cases} S_i \cup Z, & \text{when j=i} \\ S_i, & \text{otherwise} \end{cases} \tag{2.9}$$

When $i = j$, generalized class-wise statistic (T_i^j) can be updated using following simplified equation,

$$T_i^j = \frac{1}{(n_i+1)k}(n_i k T_i + \Delta n_i^j + k_i) \tag{2.10}$$

here n_i is the number of training data for class i, k_i is the number of nearest neighbor of test sample from class i, Δn_i^j represents the change of k nearest neighbor for class i when test sample is assumed to be class j and T_i represents the generalized class-wise statistic of original class i. The generalized class-wise statistics value has great significance and it represents the variability between samples in a particular class. The highest value of generalized class-wise statistics is 1, it means samples for this class are the most similar and has least variability. The less the value of generalized class-wise statistics represents higher variability between samples.

For the case of $i \neq j$, generalized class-wise statistic (T_i^j) is updated using the following equation,

$$T_i^j = T_i - \frac{\Delta n_i^j}{n_i k} \tag{2.11}$$

During successive assignment of test sample into each possible class, new-generalized class-wise statistic for other possible classes should also be determined. This means that during the consideration of test sample in class 1, new generalized class-wise statistic for classes 2, 3 .... is also determined. After getting values of all T_i^j, the ENN classifier predicts its class membership from intra-class coherence Θ^j, which is given below,

$$\begin{aligned} f_{ENN} &= \underset{j \in 1,2,\ldots}{arg\,max} \Theta^j \\ &= \underset{j \in 1,2,\ldots}{arg\,max} \sum_{i=1}^{i} T_i^j \end{aligned} \tag{2.12}$$

From this equation it is noticeable that, for determining intra-class coherenceΘ^j it is needed to compute generalized class-wise statistic.

2.4.3 Predicting Load Demand Variability and Estimating Load Demand

In our work we have utilized the generalized class-wise statistics as in (4.2), (2.11) for predicting the load demand variability. The higher the value of generalized class-wise statistics has less load variability and vice versa. Now the load demand for a customer is estimated in the following way:

If the load profile of a customer is denoted by $E(t)$, maximum load is denoted by

$E_{max}(t)$, minimum load is denoted by $E_{min}(t)$, then the estimated load demand for a particular customer is given by,

$$\hat{E}(t+1) = E_{min}(t) + T_i^j * [E_{max}(t) - E_{min}(t)] \tag{2.13}$$

here $\hat{E}(t+1)$ represents the estimated load demand of a customer.

2.5 Simulation Parameters and Performance Evaluation Criteria

In this section simulation parameters and data preprocessing for comparative study algorithms will be given first. In this work, load profile of residential and small medium enterprise (SME) customer is considered. Our task is to identify each customer from their load profile data. The load profile data were collected from Minneapolis, USA (Open EI) [66]. For comparison purpose, load profile classification accuracy and mean absolute percentage error (MAPE) will be considered as performance evaluation criteria.

2.5.1 Simulation Parameter: Neural Network

For neural network classification, normalized load profile data is considered. Here load profile includes smart meter load demand data in every hour. This smart metered load profile data are applied to neural network for classification without any attribute reduction. Back-propagation algorithm is used as learning classifier. From cross validation it is found that, for learning rate of 0.2 and maximum iteration number of 2500, classification accuracy of neural network is found to be optimum. In our considered neural network structure, there are 24 input nodes, one hidden layer with 6 nodes, and a single output node.

2.5.2 Simulation Parameter: $k-$Nearest Neighbor

During kNN classification, normalized load profile data is considered without any attribute reduction. Nearest neighbor is determined based on Euclidean distance. After cross validation it is found that, optimal classification accuracy is obtained when k equals 5.

2.5.3 Simulation Parameter: Support Vector Machine

Data normalization is done first to reduce the effect of outlier data. Then we used data dimension reduction to avoid computational complexity. In this work, after data dimension reduction of load profile, sparse signal consists of four features (1st and 2nd principle component [44], peak load demand and load demand variation [58]). Penalty parameter C plays a vital role for SVM classifier performance. From cross validation it is found that when C equals 0.2, SVM classification accuracy is optimum. Linear kernel function is used for our case.

2.5.4 Simulation Parameter: Extended Nearest Neighbor

Load profile data collected from smart meter is directly applied for classification using ENN without incurring additional computational complexity. In this case also, normalized load profile was used for classification without any attribute reduction. After cross validation, classification accuracy of ENN algorithm is found to be optimal for consideration of 5 nearest neighbor.

2.5.5 Performance Evaluation Criteria

For performance comparison among considered classification methods, classification accuracy is considered as one of the evaluation criteria [44]. Load demand prediction or estimation is evaluated based on MAPE.

1) The accuracy of the classifier is defined as the proportion of data that are correctly labeled,

$$\text{Accuracy} = \frac{\text{Total number of correct prediction}}{\text{Total number of testing data}} \tag{2.14}$$

We can determine the classification accuracy from the confusion matrix.

2) MAPE is defined as,

$$MAPE = (1/N) * \sum_{t=1}^{N} \frac{\mid E(t+1) - \hat{E}(t+1) \mid}{\mid E(t+1) \mid} * 100 \tag{2.15}$$

here $E(t+1)$ denotes actual load demand and $\hat{E}(t+1)$ denotes the predicted load demand.

2.6 Simulation Results and Analysis

Implementation of ENN for electricity load profile classification is done using Matlab R2017b on a standard PC with an Intel (R) Core (TM) i7-4790 CPU running at 2.40 GHz and 8.0 GB RAM.

2.6.1 Description of the Dataset

For classifying the electricity customers we have collected the dataset from Minneapolis, USA (Open EI) [66]. Data collection date was from 1st July 2018 to 31st December 2018 and sampling time is one hour. Load profile of 360 customers was considered for classification, half of them (i.e.,180) are residential customers and rest half are SMEs. For each of the consumer we have collected 60 load profile data. Our task is to identify individual customer from customer's load profile data. With this customer classification result we have predicted the load variability of a customer. Also, we have estimated the future load demand of a customer from previous load profile history. Typical normalized load consumption is shown in Fig. 4.3. From this figure, it is evident that individual load profile is volatile and uncertain which makes load profile classification more complex.

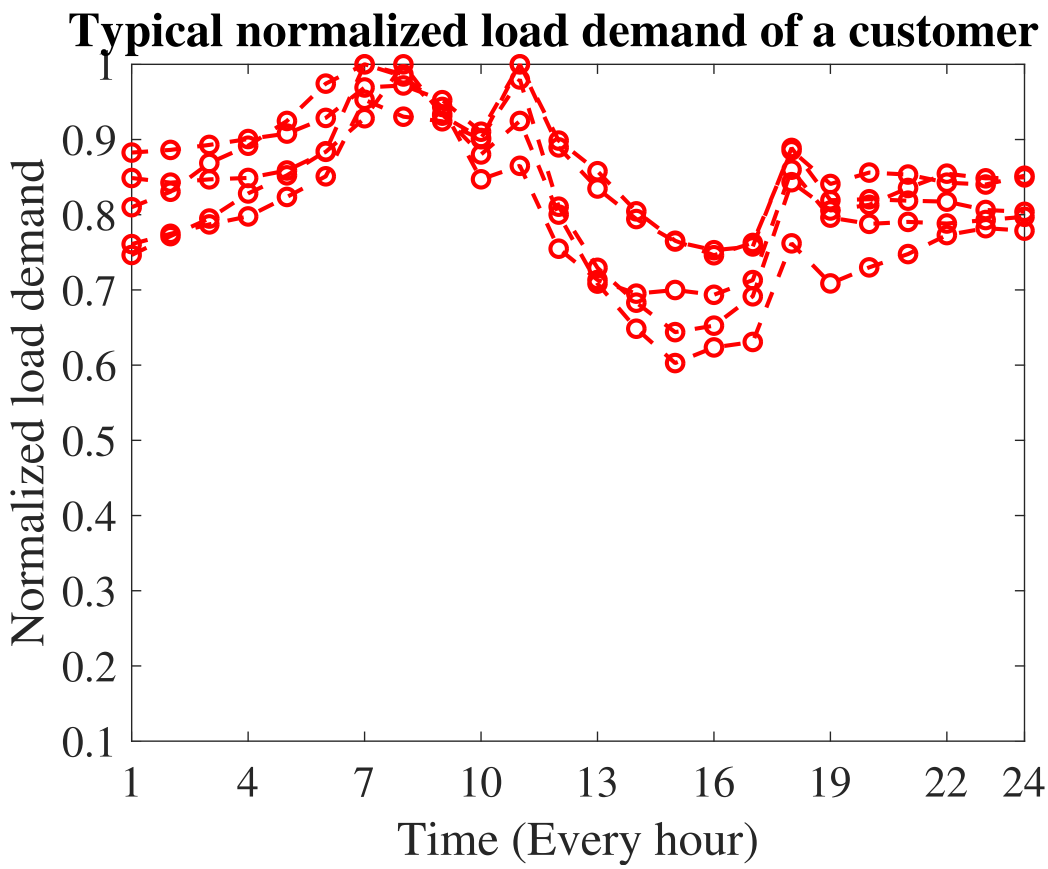

Figure 2.1. Normalized load profile

2.6.2 Experimental Results

For classification of electricity load profile dataset, half of the dataset were considered for training and rest half for testing. Training samples consists of 5400 residential load profile and 5400 SME load profile. Similarly during testing, 5400 load profile from residential customers and 5400 load profile from SME customers were considered. The training and testing samples were selected randomly to perform ENN simulation. The load profile classification accuracy is shown in Fig. 4.4. Average classification accuracy is determined after

100 simulation run. From Fig. 4.4, it is clear that load profile classification accuracy using ENN is much higher than other considered comparative algorithms. Specifically there is significant improvement in classification accuracy compared to kNN based load profile classification. The reason for performance improvement using ENN is two fold. First, ENN takes the advantages of all available training data to make classification. By exploiting the information from all available training data to maximize intra-class coherence, ENN is able to learn from global distribution, therefore improving classification accuracy. Second, ENN updates new generalized class-wise statistic in successive iteration and makes prediction for a class which gives highest generalized class-wise statistics.

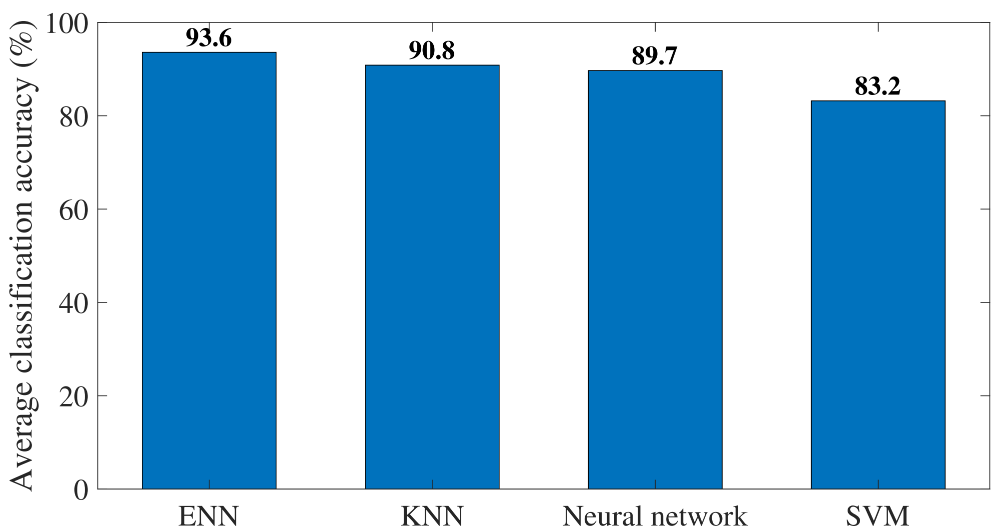

Figure 2.2. Comparison of classification accuracy among comparative algorithms

Now, let us determine the load variability of a customer. To determine the load variability of a customer we need to determine generalized class wise statistics. In this research work we have investigated 21600 load profile data of 360 customers. For convenience, we have presented the generalized class wise statistics of 10 customers as presented in Table 4.2. The less the value of the generalized class wise statistics, the large the load demand variability for that customer. The large variability of load demand is not good. If the generalized class wise statistics is 1, it means load variability for that customer is minimum. As seen in Table 4.2, customer #5, #6, #7, #8, #9 has highest generalized class wise statistics i.e., 1, therefore these customers has minimum load demand variability. Demand response programs can target these customer for energy efficiency improvement. These customer are suitable for determining operational planning of demand response programs. The customer #1 has lowest generalized class wise statistics i.e., 0.30 which has the highest load demand variability and it can vary more than double of usual load demand.

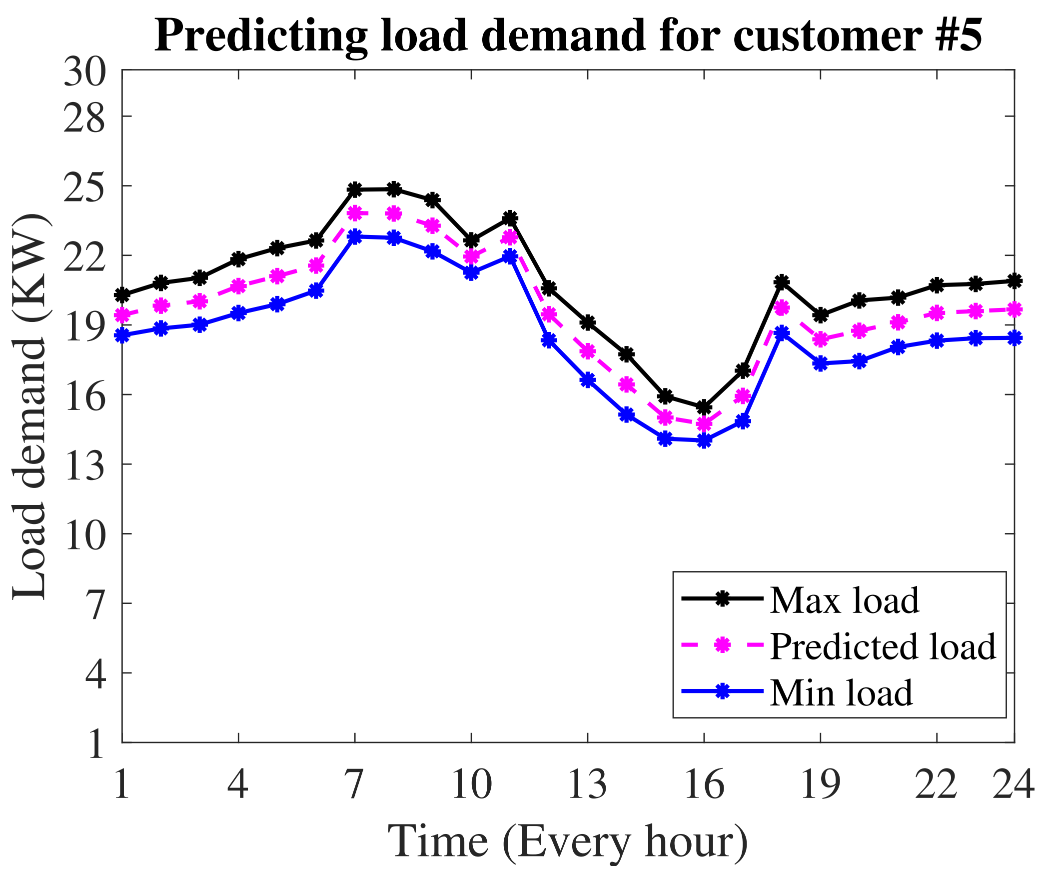

Figure 2.3. Load demand prediction of customer #5

The prediction of load demand for a particular customer 5 is shown in Fig. 4.6, and load demand is estimated from generalized class wise statistics, maximum load demand, minimum load demand. The estimated load demand is determined using (2.13). The accuracy of this load demand estimation is evaluated based on MAPE value. The less the MAPE value means higher the prediction accuracy. The value of MAPE is dependent on generalized class wise statistics and load demand variability of a customer. The MAPE value is 4.98% which is in permissible limit.

Table 2.1. Predicting load demand variability.

Customer Number #	Generalized Class-wise Statistics	Customer Number #	Generalized Class-wise Statistics
1	0.30	6	1.00
2	0.45	7	1.00
3	0.58	8	1.00
4	0.31	9	1.00
5	1.00	10	0.32

2.7 Summary

In this chapter, customers are classified from electricity load profile data using a recently developed classification algorithm called ENN. The advantage of this classification approach is: it provides the idea of load demand variability of a customer and helps to estimate the load demand of a particular customer using the classification result. The challenge of fair accuracy in case of kNN algorithm is suppressed after using ENN algorithm. The results indicates that, the proposed ENN classifier achieves higher accuracy. Different from traditional load profile classification approaches which mainly focuses on the shape of load profile or data dimension reduction techniques, this paper tries to perform load profile classification directly from smart metered data i.e., removing preprocessing stage. Moreover, ENN classifier learns from global distribution of all available training samples for predicting the class of an unknown test sample. The load profile classification result can be used to model the electricity consumption dynamic characteristics of each customer.

CHAPTER 3 A New Composite Model for Short-term Load Forecasting

3.1 Overview

Day-ahead load forecasting is an important task for the reduction of electricity waste and efficient management of a smart grid. The electricity load profile data reveals the correlation of electricity load demand with weather condition, day type (working day or holiday), time of the day and season of the year. Thus the load forecasting problem has a high degree of complexity with consideration of those variables as input. To solve the problem of day-ahead short-term load forecasting (STLF), the proposed solution first classifies load profile data into different classes. To this end, a recent developed classification approach called extended nearest neighbor (ENN) algorithm is adopted. Then, a composite ENN model is proposed for day-ahead load forecasting. The composite ENN model consists of three individual ENN models which are combined together by tuned weight factors for predicting final forecasting output. By exploiting intra-class coherence from the generalized class-wise statistic of all available training samples, the composite ENN algorithm is able to learn from global distribution and therefore improve the accuracy of load forecasting. The proposed method is validated on two case study: (i) Australian National Energy Market Data and (ii) Brookings, South Dakota, USA Data. For case study 1, mean absolute percent error (MAPE) of composite ENN based load forecasting is decreased by 44.68% compared to composite kNN based load forecasting and mean absolute error (MAE) is decreased by 45.52%. Similarly for case study 2, the decrease of MAPE and MAE values are 27.72% and 31.65% respectively.

3.2 Introduction

Energy management systems are designed to monitor, optimize, and control the smart grid energy market. Demand-side management, considered as an essential part of the energy management system, can enable utility market operators to make better management decisions for energy trading between consumers and the operator [1], [2]. In this system, a priori knowledge about the energy load pattern (e.g., day-ahead forecasted load) can help reshape the load and cut the energy demand curve, thus allowing a better management and distribution of the energy in smart grid energy systems. Designing a computationally intelligent load forecasting system is often a primary goal of energy demand management. The accurate electricity load forecasting has a significant role in power system. It is useful for making optimal decisions to ensure the secure, reliable and economic operation of the power system [3]–[5].

The problem of STLF can be viewed as a time-series prediction problem (e.g.,[9]) in which the load is predicted based on the current-day load. Moreover, electricity load demand depends on several factors including weather, time, and socio-economic constraints [10]. Not considering the date, temperature, and other weather influences, such models can produce poor forecasting performance. When heterogeneous exogenous variables are considered as input for the load forecasting model, the STLF problem becomes complex. The STLF problem has been addressed with different methods in the literature such as regression [11], fuzzy logic [12], artificial neural network (ANN) [13]–[15] and exponential smoothing methods [16]. Each has many different models, for example, the regression based method includes auto-regressive moving average [17], autoregressive integrated moving average [18], and support vector regression [19]. Likewise, ANN based method includes bagged ANN, cascaded ANN [20], radial basis functions neural networks (RBFNN) [21], back propagation neural network (BPNN) [22] and extreme learning machine (ELM) [23]. ANN based models are the most popular methods. Rui Zhang et.al., [23] combined a series of ANN models for developing the ELM to further improve the load forecasting accuracy. Neural network based models have been used widely in load forecasting due to their learning capability of complex nonlinear relationships between load demand and the effects of historical data [24]. However, neural network based models are associated with some potential drawbacks: overfitting of the model, sensitivity to random weight initialization, and tendency to converge toward local optima [25]. Later, in 2016 Rui Zhang et.al., [26] proposed a composite k-nearest neighbor (kNN) model for day-ahead load forecasting with temperature forecasts. Authors have provided a computationally efficient & simple model compared to their previous work [23]. But it is noteworthy that neighbor selection based on kNN is affected by irrelevant features which affects the load forecasting accuracy. In data science research, to address the problems of kNN based models, B.Tang et.al., [45] proposed a novel algorithm called extended nearest neighbor (ENN) for classification application. Load profile classification from ENN algorithm is not directly affected by irrelevant features. Partitioning load profile data into groups using ENN provide generalization to yield important characteristics within load groups [67]. Precise knowledge of load profile classification will provide opportunities for accurate day-ahead load forecasting [47].

Motivating from works in [23], [25], [26], we proposed a novel load forecasting model in this project. To solve the STLF problem and further improve the forecasting accuracy, the proposed solution first classifies the load profile data into different classes according to various seasons of the year. The ENN is incorporated to solve the problem of kNN based models. The proposed composite model consists of three different individual models and the result of the models are combined together by tuned weight factor for making a final forecasting output. The contribution of this project is to develop a composite ENN model for load forecasting.

3.3 Background of Load Forecasting

If the load profile for a day m is defined as $L_m(t) = [L_m(1), ..., L_m(N)]^T$, where $L_m(t)$ is electricity load demand on mth day at $t = 1, 2, 3, ..., N$th time interval. The task associated with STLF model is to predict the load demand $\hat{L_{m+1}}(t)$ of the next day at once for the purpose of the day-ahead electricity energy market.

$$\hat{L_{m+1}}(t) = \sum_{i=1}^{P} L_{m+1}^{i}(t) \tag{3.1}$$

Here P represents the total number of models to represent the non-linear relationship between load and historical data. From this equation, final forecasted load $\hat{L_{m+1}}(t)$ is predicted based on the output from different models $L_{m+1}^{i}(t)$. Different models predict's load based on different input attributes. Input attributes are: system load demand, day type, day index, weather data etc. For our case we have used three models to predict load from system load, day type, day index and temperature data. There are different ways to predict the load from model output. In [26], Rui Zhang et.al., employed composite kNN model for day-ahead load forecasting where basic principle is to pick up k nearest samples and uses the average value of them as the forecasting output. But the nearest neighbor sample selection through kNN with considering heterogenous input exogenous variable negatively affects the accuracy of forecasting [23], [26].

Data Preparation: The analysis of load demand time series reveals significant differences among daily load profiles between weekends, weekdays, and seasons of the year [43]. To include this findings into the model, first special date variable $D^*(m)$ is introduced to take into consideration of day type of m. Variable $D^*(m)$ has value 0 at weekdays (W_0) and 1 at weekends (W_1) & holidays H:

$$D^*(m) = \begin{cases} 1, & \text{if} \quad m \in (W_1, H) \\ 0, & \text{if} \quad m \in W_0 \end{cases} \tag{3.2}$$

here W_0, W_1 and H represents the set of working day, weekend and holiday date variables.

3.4 Extended Nearest Neighbor Algorithm

3.4.1 Framework of ENN

ENN algorithm is a new supervised classification approach that predicts class label from the maximum gain of intra-class coherence [45]. It makes a prediction in a two-way communication style: it considers not only the nearest neighbors of test samples but also who (training samples) consider the test sample as their nearest neighbor. This is different from other nonparametric approach because it does not consider only the nearest neighbor for class prediction, rather makes class prediction from the maximum gain of intra-class coherence as shown in Fig. 4.1.

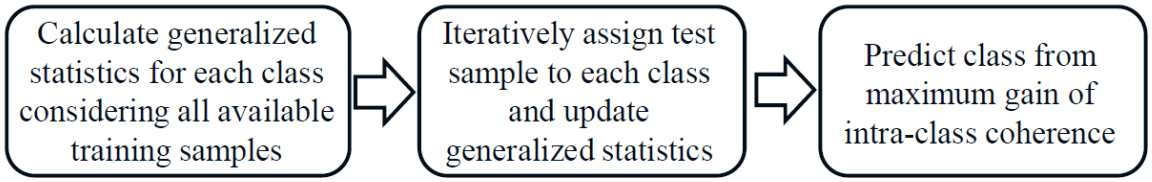

Figure 3.1. Framework of ENN algorithm for classification.

3.4.2 Recap of Key Formulas of ENN

In the ENN algorithm, the first step is to determine the generalized class-wise statistic. It is determined from the global distribution of all available training samples, which gives the idea about the distribution of dataset, i.e., whether they are mixed well or widely separated. A large value of the generalized class-wise statistic means samples for that particular class are closer together, whereas a smaller value indicates samples of that class are not closer together. Note that, it's value lies in the range of 0 to 1. For each class of data samples the generalized class-wise statistic T_i for class i is determined as [45],

$$T_i = \frac{1}{n_i k} \sum_{x \in S_i} \sum_{r=1}^{k} I_r(x, \Re = (S_1 \cup S_2 \cup \ldots \cup S_M)) \tag{3.3}$$

where n_i represents the total number of samples in S_i, class label is represented by i, S_1, S_2,... S_M, denotes set of corresponding samples in each of the class, M denotes the total number of class labels or the total number of group, x denotes one single sample in S_i and k is user defined parameter which represents the number of nearest neighbors. Nearest neighbors are determined based on euclidian distance measurement. In this equation indicator function $I_r(x, \Re)$ is used to check whether both the sample x and its $r-th$ nearest neighbor belongs to the same class or not. A large T_i indicates the samples in S_i are much closer together and their nearest neighbors are dominated by the same class samples, whereas a small T_i indicates that samples in S_i have an excess of nearest neighbors from the other class. Note that the generalized class-wise statistic has $0 \leq T_i \leq 1$ with $T_i = 1$ when all the nearest neighbors of class i data are also from the same class i , and $T_i = 0$ when all the nearest neighbors are from other classes. Based on this discussion, we can use T_i to represent the data distribution across multiple classes. Therefore, we will introduce the concept of intra-class coherence Θ at later to predict the class label.

Indicator function $I_r(x, \Re)$ is defined as,

$$I_r(x, \Re) = \begin{cases} 1, & \text{if} \quad x \in S_i \quad \text{and} \quad \Re \in S_i \\ 0, & \text{otherwise} \end{cases} \tag{3.4}$$

where $\Re$ represents $r-th$ nearest neighbor of sample x in S_i. The outcome of indicator function is 1 if both the sample x and it's $r-th$ nearest neighbor belongs to the same class. Conversely for dissimilarity between the sample and it's nearest neighbor's class, indicator function outcome is zero.

In the testing phase, the unknown test sample is iteratively assigned to each possible class. During each iteration, the generalized class-wise statistic is updated. For updating new generalized class-wise statistic T_i^j, the following equation is considered,

$$T_i^j = \frac{1}{n_i' k} \sum_{x \in S_{i,j}'} \sum_{r=1}^{k} I_r(x, \Re' = (S_1 \cup S_2 \ldots S_i \cup Z)) \tag{3.5}$$

Here Z represents the new set of a sample when it is assigned in a new class during the successive iteration process, n'_i is the size of $S'_{i,j}$, and $S'_{i,j}$ is defined as follows,

$$S'_{i,j} = \begin{cases} S_i \cup Z, & \text{when j=i} \\ S_i, & \text{otherwise} \end{cases} \tag{3.6}$$

When $i = j$, the generalized class-wise statistic (T_i^j) can be computed from the following simplified equation,

$$T_i^j = \frac{1}{(n_i+1)k}(n_i k T_i + \Delta n_i^j + k_i) \tag{3.7}$$

here n_i is the number of training samples in class i, k_i is the number of nearest neighbor of the sample from class i, Δn_i^j represents the change in the number of nearest neighbors for class i when test sample is assumed to be class j and T_i represents the generalized class-wise statistic of original class i. Similarly for the case of $i \neq j$, generalized class-wise statistic (T_i^j) can be calculated from following simplified equation,

$$T_i^j = T_i - \frac{\Delta n_i^j}{n_i k} \tag{3.8}$$

During successive assignment of the test sample into each possible class, the new-generalized class-wise statistic for other possible classes should also be determined. The ENN classification algorithm predicts class membership of an unknown test sample from following equation,

$$\begin{aligned} f_{ENN} &= \underset{j \in 1,2,\ldots}{arg\,max} \Theta^j \\ &= \underset{j \in 1,2,\ldots}{arg\,max} \sum_{i=1}^{M} T_i^j \end{aligned} \tag{3.9}$$

Note that, after getting values of all T_i^j, ENN algorithm determines maximum gain of intra-class coherence Θ^j.

3.5 Design of Composite ENN Model for STLF

3.5.1 Three Individual Developed Models

1) Model-I: In this model, we have considered only the day-type $D^*(m)$ and neglected the temperature impacts. To forecast the load demand of day $(m+1)$, first load profile is classified based on it's load demand and day type $D^*(m)$. Then it picks up load demand of k days which has the same class label and day type. The forecasting output is the average of the load values of the k days, i.e.,

$$L^1_{m+1}(t) = (1/k) * \sum_{m=1}^{k} L_m(t)(D^*(m)) \tag{3.10}$$

2) Model-II: In the second model, temperature data is considered with day type. Similar to Model-I, classification is done first to get label and generalized class-wise statistics. Then it picks up load demand of k days which has the same day type to forecast load demand of $(m+1)$th day. The forecasting output is the average of the load values of the k days, i.e.,

$$L^2_{m+1}(t) = (1/k) * \sum_{m=1}^{k} L_m(t)(\theta_m, (D^*(m))) \tag{3.11}$$

3) Model-III: The third model takes the day-index and the temperature data as input. To forecast load demand of day $(m+1)$th day, classification is done first as was done for other models. If the day-index is denoted as $\rho(m)$ and number of nearest neighbors are k days, then the forecasting output is the average of the load values of the k days, i.e.,

$$L^3_{m+1}(t) = (1/k) * \sum_{m=1}^{k} L_m(t)(\theta_m, \rho(m)) \tag{3.12}$$

3.5.2 Ensemble Strategy

In recent times, researchers have followed the ensemble strategy as an effective means to increase the accuracy of a single model. The philosophy of ensemble strategy is to combine a series of single models to make a final prediction. For regression problem, the final forecasting value is made from the average value of individual outputs [68]. In doing so, single learners can compensate for each other, and the whole can reduce aggregated variance and tend to increase the accuracy over the individuals.

Following the strategy applied in [26], in this paper we have proposed a composite ENN model for load forecasting. Since a single model is inadequate to capture the inherent complexity of the time series, thus we propose to use a composite model consisting of Model-I, Model-II, and Model-III. The diversity among these models lies in the input applied to models and the method of selecting nearest neighbor samples. The output from these three models are aggregated through tuned weights obtained from generalized class-wise statistics.

$$\hat{L_{m+1}}(t) = \omega_1 * L^1_{m+1}(t) + \omega_2 * L^2_{m+1}(t) + \omega_3 * L^3_{m+1}(t) \tag{3.13}$$

here $\hat{L_{m+1}}(t)$ denotes the final forecasted load of day $(m+1)$ from 1st to the N-th time horizon, $L^1_{m+1}(t)$, $L^2_{m+1}(t)$, and $L^3_{m+1}(t)$ represents the output obtained from Model-I, Model-II, and Model-III respectively. And ω_1, ω_2, ω_3 are the weighting factors for three models.

3.5.3 Performance Evaluation Criteria

For evaluating load profile classification performance, the classification accuracy will be considered as the only criteria. Then the performance of load forecasting models are compared with respect to mean absolute percentage error (MAPE) and mean absolute error

(MAE) [23], [26].

1) The accuracy of the classifier is defined as the proportion of data that are correctly predicted [59],

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.14}$$

From the confusion matrix given below, we can obtain the number of true positive (TP), false negative (FN), false positive (FP) and true negative (TN).

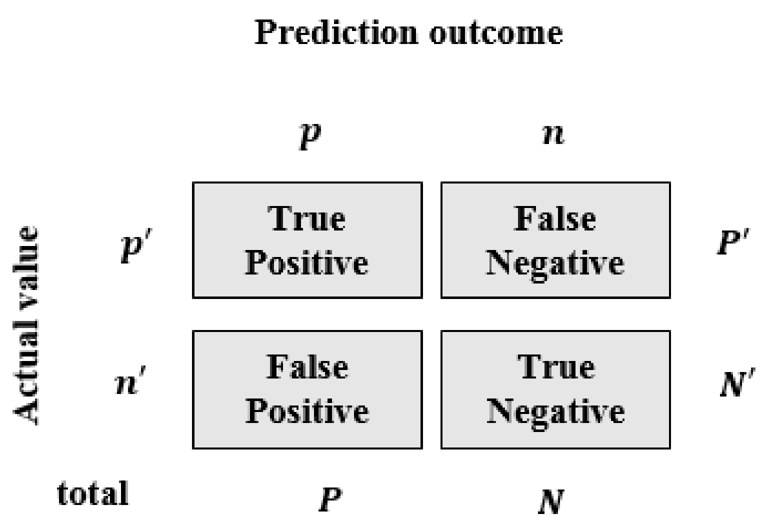

Figure 3.2. Confusion matrix.

2) MAPE is defined as,

$$MAPE = (1/N) * \sum_{t=1}^{N} \frac{\mid L_{m+1}(t) - \hat{L}_{m+1}(t) \mid}{\mid L_{m+1}(t) \mid} * 100 \tag{3.15}$$

here $L_{m+1}(t)$ denotes actual load demand and $\hat{L}_{m+1}(t)$ denotes the forecasted load demand.

3) MAE is defined as,

$$MAE = (1/N) * \sum_{t=1}^{N} \mid L_{m+1}(t) - \hat{L}_{m+1}(t) \mid \tag{3.16}$$

3.6 Simulation Results and Analysis

3.6.1 Description of the Dataset

For case study 1, the data is collected from 1st January 2009 to 31st December 2010 with sampling time of half hour for the region of New South Wales (NSW) , Australia [69], [70] provided by Australian National Energy Market (NEM) Operator. This means, for every day there are 48 data samples for load demand time series. Here this dataset is classified into four groups to represent four seasons of the year. The plot of load variation in different seasons of the year for case study 1 is shown in Fig. 3.3. This case study is for commercial customer. With different seasons, the load demand changes significantly in terms of their value which is clearly evident from Fig. 3.3. For the same time span, we have also collected the dataset for temperature and it varies from 5.8 degree Celsius to 43.8 degree Celsius as shown in Fig. 3.4. We assumed that we have information of holidays, weekends within a year. This variable has value 0 at weekdays, and 1 at weekends & holidays.

For case study 2, the data is collected from 1st January 2017 to 31st December 2018 with sampling time of one hour for the residential customers of Brookings, South Dakota, USA [71]. This means, for every day there are 24 data samples for load demand time series. This dataset is for residential customer which includes 124 homes. For each of the customer, we have collected three main groups of measured variables: weather data (temperature), time categorical data (hour, month, day), and electrical load. Likewise, for this case study also the dataset is classified into four groups to represent the four seasons of the year. The plot of load variation in different seasons of the year for case study 2

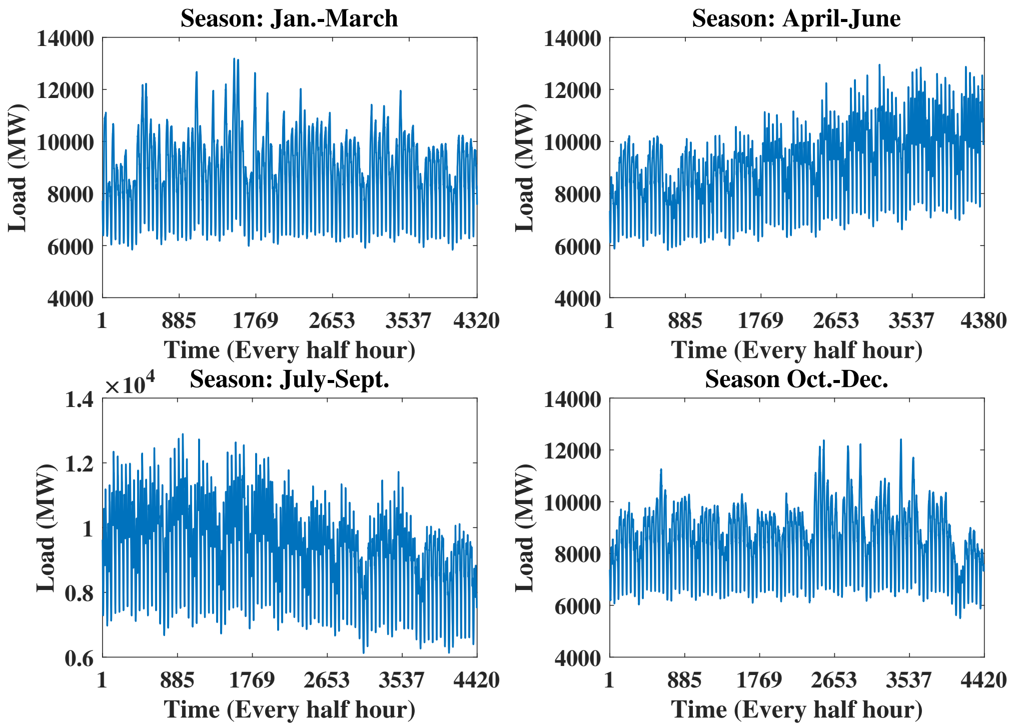

Figure 3.3. Load data used for Case study 1: Location- New South Wales (NSW), Australia.

is shown in Fig. 3.5. Consistently, in this case study also, with different season the load demand changes significantly in terms of their value which is clearly evident from Fig. 3.5. For same time span, we have also collected the dataset for temperature and it varies from -6.8 degree Celsius to 24.2 degree Celsius as shown in Fig. 3.6.

3.6.2 Selection of k and Weighting Factors

All the simulations are conducted using Matlab R2017b on a standard PC. In line with the tests conducted in [23], [26], the data of the year 2009 is used for training the composite ENN model. In the training phase, the value of k and weighting factors ω_1, ω_2, ω_3 are tuned. The data of case study 1 for the year 2010 is used to test the generalization performance of the tuned model. The validation MAPE of the three ENN models at different values of k are given in Fig. 4.3.

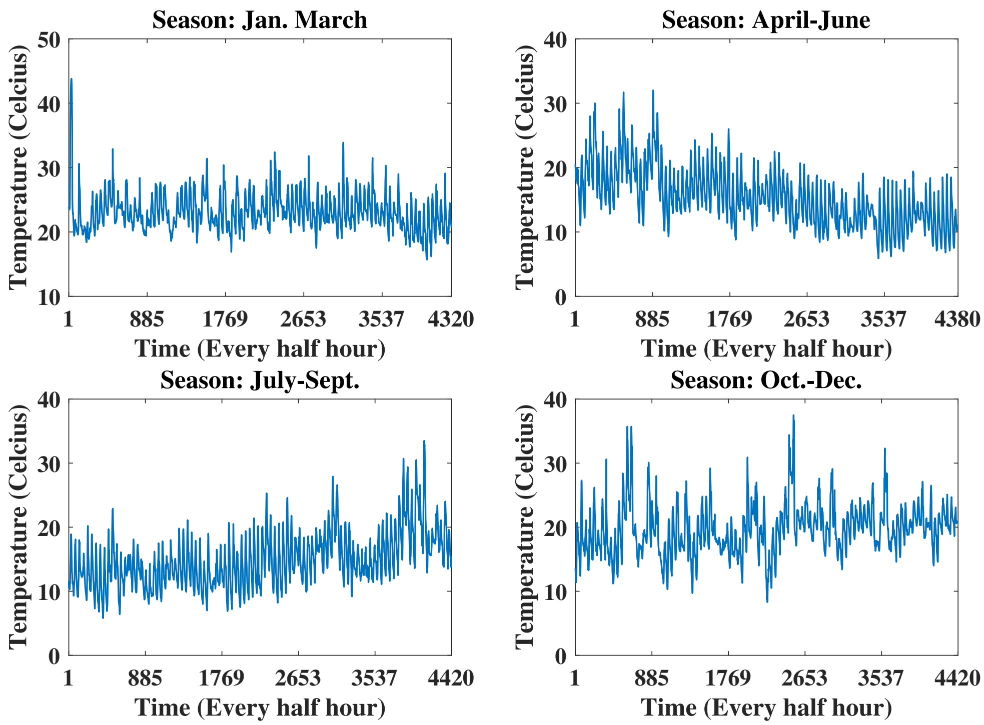

Figure 3.4. Temperature used for Case study 1: Location- New South Wales (NSW), Australia.

According to Fig. 4.3, it is seen that initially when k is 1, the forecasting error of the three models are relatively big (i.e., MAPE values are 6.56%, 7.23%, and 8.16% for Model-I, Model-II, and Model-III, respectively). Theoretically, this should be the weak learner to use for composite ENN model. Yet, from our study, this setting doesn't provide the best ensemble learning performance and so we use the k with minimum MAPE error. From Fig. 4.3, it is found that if $k = 5$, the error values of MAPE are 2.15%, 2.68%, and 2.86% for Model-I, Model-II, and Model-III, respectively. When load vector is considered for classification it is found that among 5 nearest samples, 3 come from same class and generalized class-wise statistics indicates 0.6 variability. So the weighting factor ω_1 is taken as 0.6. This weight parameter is tuned during the calculation of generalized class-wise statistics from equation (3.3). Similarly, the other two factors indicate variability of

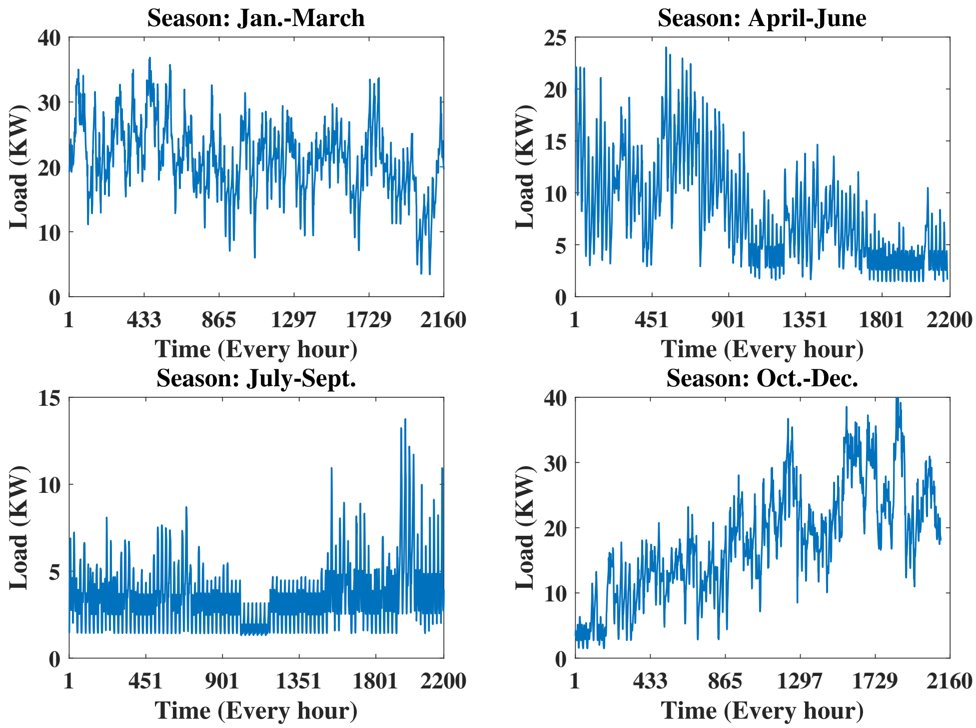

Figure 3.5. Load data used for Case study 2: Location- Brookings, South Dakota (SD), USA.

0.2 for each models and hence ω_2, ω_3 are 0.2. During this time, equations (3.7) & (3.8) are considered. Instead of using empirical methods from [26], we tune the weighting factors based on the changes in generalized class-wise statistics due to consideration of two-way communication between training and testing samples.

3.6.3 Experimental Results

Case study 1: Australian NEM data are investigated firstly for load profile classification and then for load forecasting. Initially the dataset is divided into four classes for four seasons of the year. This classification of dataset permits generalization and will reduce the influence of irrelevant features. In this stage the label of load data determination provides generalization over a class. Later during forecasting stage this classification is employed

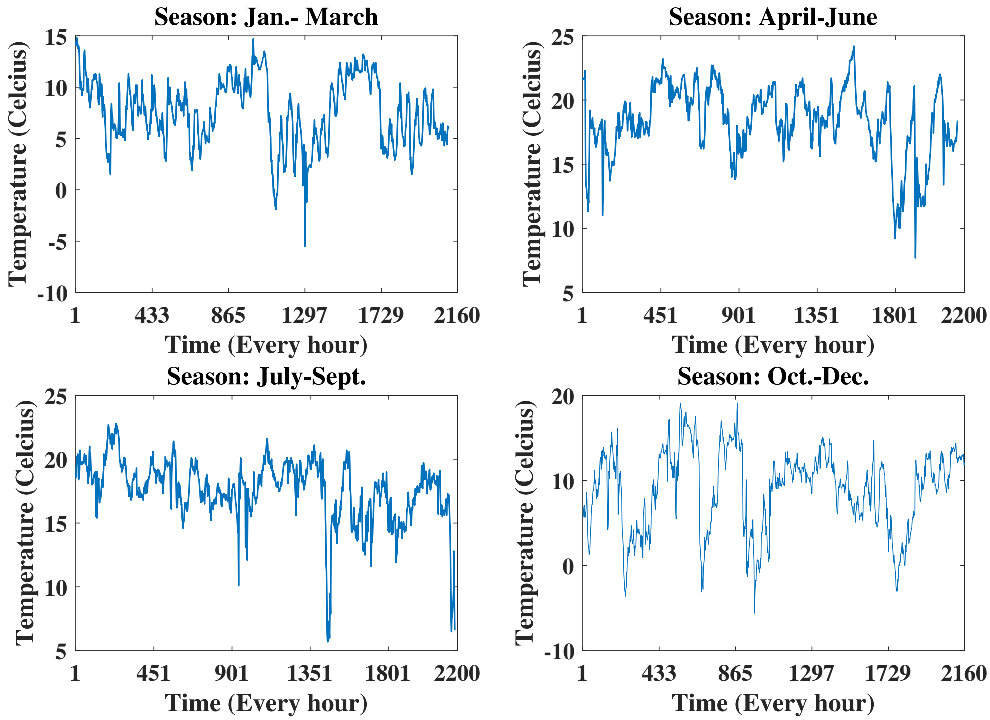

Figure 3.6. Temperature data used for Case study 2: Location- Brookings, South Dakota (SD), USA.

to identify the label of nearest neighbor samples. The performance of ENN classifier is presented in TABLE 4.2. The higher classification accuracy demonstrates the effectiveness of ENN classifier. At the same time generalized class-wise statistics is determined from equations (3.3), (3.7) and (3.8).

Table 3.1. Load Profile Classification Result Using ENN for Case Study 1, Location - NSW, Australia.

Season	Tr. Data (days)	Test. Data (days)	Tr. Accuracy (%)	Testing Accuracy (%)
Jan.-March	90	90	98.88	97.77
April-June	91	91	100	98.90
July-Sept.	92	92	100	98.01
Oct.-Dec.	92	92	97.82	94.74

After predicting the class label of load data, the next step is forecasting stage. For load forecasting, 5 nearest neighbor samples are picked with the help of load profile clas-

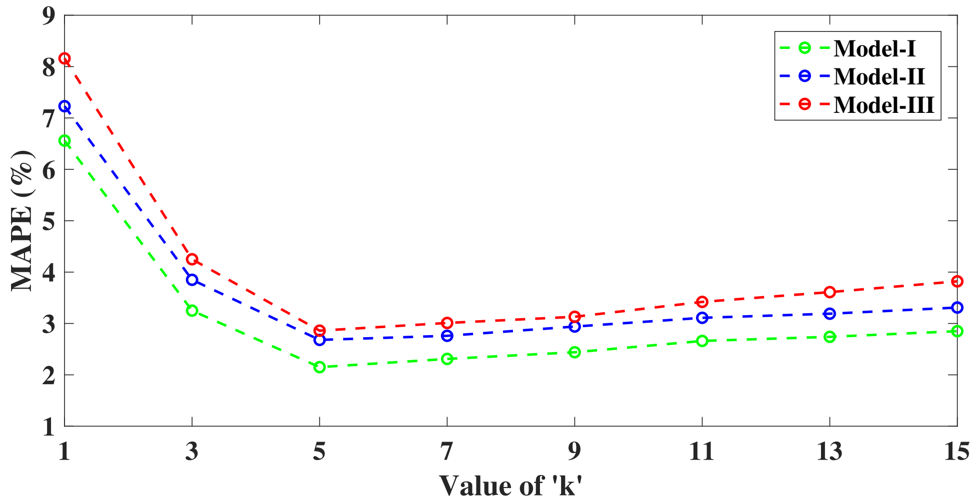

Figure 3.7. Validation MAPE with different value of k.

sification. Load demand of these nearest neighbor samples are combined and average of them predicts next-day load demand. The resulting forecasted load from proposed model is shown in Fig. 4.4. The provided simulation was conducted for a day in Jan.-March of the year 2010. The performance from the proposed model is compared with the other five models presented in [23], [26], which are composite kNN model, BPNN, RBFNN, ELM and ensemble ELM. The BPNN and RBFNN are popular methods for load forecasting and ELM is an emerging ANN learning algorithm. MAPE and MAE values of different forecasting methods are presented in TABLE 4.3. From the simulation results it is apparent that, proposed composite ENN model provides better performance than other comparative algorithms. Compared with the results presented in [23], our method yields MAPE value of 1.77% and MAE value of 159.02 MW. With reference to the result presented in [26], the performance of the proposed method also decreases MAPE by 44.68% and MAE by 45.52%. Now if we compare the performance of individual model with forecasted load, we got that MAPE value of Model-I, Model-II, and Model-III are 2.34, 2.18, and 2.10 respectively. However, the composite ENN model yields MAPE value of 1.77. Likewise three individual model yields MAE value of 228.98, 212.54, and 198.48, whereas the composite ENN model yields MAE value of 159.02.

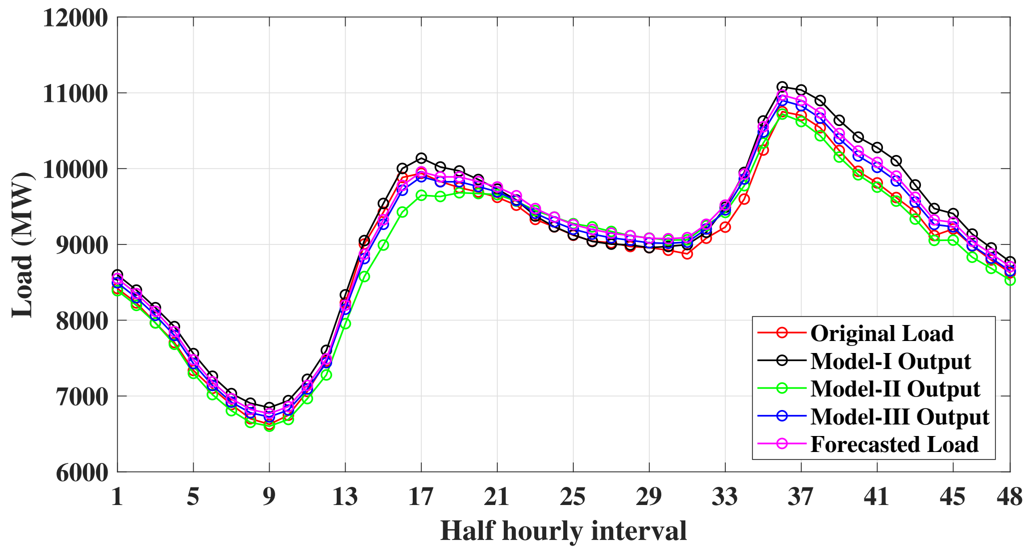

Figure 3.8. Load forecasting with proposed model of Case study 1 for different regions of Australia.

Table 3.2. Load Forecasting Performance Comparison for Case study 1.

Model	MAPE (%)	MAE (MW)
Composite ENN	1.77	159.02
Composite kNN [26]	3.20	291.9
ELM [23]	2.89	260.6
BPNN [23]	2.93	266.2
RBFNN [23]	2.86	258.1
ELM ensemble [23]	1.82	171.7

Case study 2: For this case study we have collected data set of Brookings, South Dakota, USA [71]. During this case study the data is collected at every hour from residential customer. In a similar manner as in case study 1 we have carried out the simulation with the same parameter setting. In this case we have considered individual customer and for each customer we have collected the data set for two year. Likewise, we have considered four seasons in a year. Here, for each individual customer we have done load profile classification for four seasons of the year. Likewise of case study 1, we have performed the load profile classification first and the classification result is provided in TABLE 3.3. From this classification result it is found that, ENN can classify load profiles with higher classification accuracy.

Table 3.3. Load Profile Classification Result Using ENN for Case Study 2, Location-Brookings, SD, USA.

Season	Tr. Data (days)	Test. Data (days)	Tr. Accuracy (%)	Testing Accuracy (%)
Jan.-March	90	90	100	98.88
April-June	91	91	98.90	97.80
July-Sept.	92	92	100	98.91
Oct.-Dec.	92	92	98.91	97.82

The forecasted load profile is shown in Fig. 4.2. MAPE and MAE value of different forecasting methods are presented in TABLE 3.4. Comparison among the different methods presented here shows that, our proposed method yields better performance than the others. If we compare the performance of our proposed method with the previous work [26], there is 27.72% decrease in MAPE value and 31.65% decrease in MAE value compared to composite kNN model based load forecasting. In this case, if we compare the performance of individual models with forecasted load, we got that MAPE values of Model-I, Model-II, and Model-III are 13.76, 10.97, and 10.36 respectively. However, the composite ENN model yields MAPE value of 2.79. Likewise three individual models yield MAE values of 286.10, 242.57, and 214.61, whereas the composite ENN model yields MAE value of 63.00. From this case study result, the advantage of employing composite model instead of single individual model is evident. In this case study it is apparent that, proposed composite

ENN model provides better performance than other comparative algorithms.

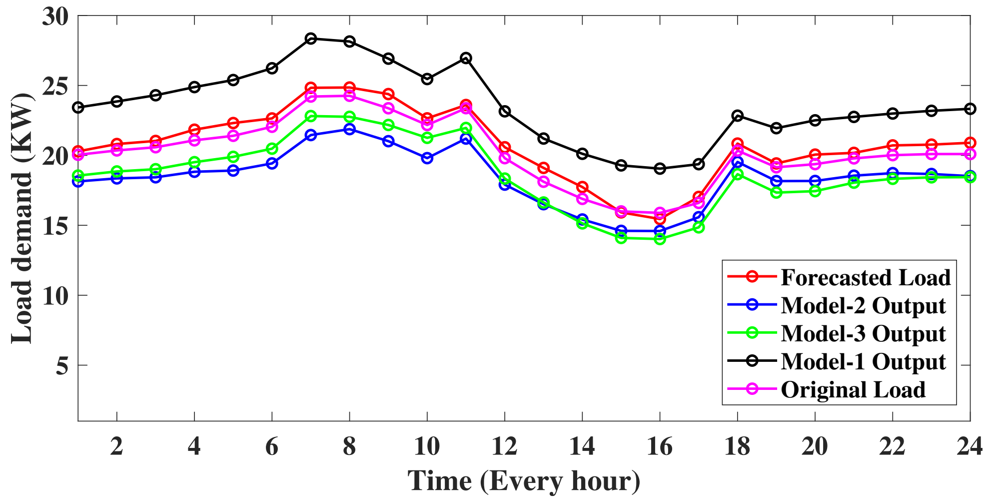

Figure 3.9. Load forecasting with proposed model for Case study 2 of Location- Brookings, South Dakota, USA.

Table 3.4. Load Forecasting Performance Comparison for Case study 2.

Model	MAPE (%)	MAE (KW)
Composite ENN	2.79	63.00
Composite kNN [26]	3.86	92.18
ELM [23]	3.94	98.21
BPNN [23]	4.24	112.48
RBFNN [23]	3.68	82.28
ELM ensemble [23]	3.16	76.44

The reason for performance improvement using ENN algorithm is two-fold. First, ENN algorithm exploits the information from all available training samples to maximize the gain of intra-class coherence, which is learned from global distribution. The presence of the negative term in equation (7) means that, a class with large generalized class-wise statistic would be given a higher penalty value for the wrong estimation of class membership of an unknown testing sample. Moreover two-way communication between testing and training samples removes the influence of irrelevant features. Load profile classification facilitates generalization which improves the forecasting accuracy. In addition it is noteworthy that, the application of composite model also improves the load forecasting accuracy due to it's ability to extract information from different variety of input.

3.7 Summary

In this project, the solution of the STLF problem is presented by separating the load profiles into different classes with the ENN algorithm. The developed composite model predicts the next-day's load considering exogenous input variables. Composite model consists of three different individual models designed with different varieties of input attributes to incorporate the inherent complexities with load demand time series data. The three individual models are combined together by designated weighting factors to provide a final forecasting output. The developed model is validated on the Australian National Energy Market data, & Brookings Data and compared with the results reported in previous studies. The case study results verify that, the proposed composite model with ENN load profile classification provides higher accuracy. In addition to accuracy improvement, the proposed composite model reduces the effect of irrelevant features due to load profile classification based on generalized class wise statistics. Specifically the determination of generalized class wise statistics by two-way communication removes the effect of irrelevant feature in load forecasting accuracy.

CHAPTER 4 A New Hybrid Model for Short-term Load Forecasting

4.1 Overview

Prior knowledge of electricity load demand i.e., load forecasting can help utility operators for the efficient management of a demand response program. Forecasting electricity load demand with higher accuracy and efficiency is a challenging task since electricity load is affected by previous history load, several exogenous external factors (i.e., weather variables, social variables, working day or holiday), time of day, and season of the year. To solve the problem of short-term load forecasting (STLF) and further improve the forecasting accuracy, in this chapter we have proposed a novel hybrid STLF model with new signal decomposition and correlation analysis technique. To this end, load demand time series are decomposed into some regular low frequency components using novel improved empirical mode decomposition (IEMD). To compensate for the information loss during signal decomposition, we have incorporated the effect of exogenous variables by performing correlation analysis using T-Copula. From the T-Copula analysis, peak load indicative binary variable is derived from a new parameter i.e., value at rsik (VaR) to improve the load forecasting accuracy during peak time. The output obtained from IEMD and T-Copula is applied to deep belief network for predicting the future load demand of specific time. The proposed data driven method is validated on real time data from the Australia and the United States of America. The performance of proposed load forecasting model is evaluated in terms of mean absolute percentage error (MAPE) & root mean square error (RMSE).

4.2 Introduction

Over the last few years, researchers have proposed many models to forecast electricity load for varying time interval. Based on the model architecture, load forecasting models are primarily divided into two classes: traditional statistical models and advanced data driven models. Traditional statistical models are built using linear regression function where the problem of STLF is viewed as a time-series prediction problem [9], [11]. The regression based models include auto-regressive moving average [17], autoregressive integrated moving average [18], autoregressive moving average with exogenous variable [27] and support vector regression [19]. The regression based models are effective for predicting stationary time series. However, load demand time series is non-stationary and shows nonlinear characteristics, thus advanced data driven models are proposed in recent times. To date, STLF problem has been investigated with different advanced data driven models. Advanced data driven models include: fuzzy logic based [12], artificial neural network (ANN) based [13], [14] and exponential smoothing methods [16]. ANN based models are the most popular among advanced data driven models. The ANN based method includes: bagged ANN, cascaded ANN [20], radical basis functions neural networks [21], back propagation neural network [22] and extreme learning machine [23]. Both statistical and advanced individual data driven models are proposed to predict the load demand. However, a single model is inadequate to represent inherent characteristics of electricity load demand because it depends on several factors including weather, time, and socio-economic constraints [10]. If we do not consider the date, temperature, and other weather influences, such models produce fair forecasting performance. When heterogeneous external factors are considered as input for the load forecasting model, the STLF problem becomes complex.

Thus, hybrid models are formed by integrating different models for improving the forecasting accuracy. The reason is that, different models can capture the features of electricity load profiles. In general, the hybrid models are classified into two main categories. For the first category model, electricity load is predicted separately by different models [28]–[33]. For the second category model, electricity load is decomposed into several components. Then each component is predicted by a suitable model [34]–[40]. Motivating from the works in [28]–[40], in this chapter we will present our proposed novel hybrid load forecasting model which includes new signal decomposition technique and new correlation analysis technique. To mitigate end effect and envelope fitting limitation associated with traditional empirical mode decomposition (EMD), a new improved empirical mode decomposition (IEMD) method is proposed. By using IEMD, the original load demand time series is decomposed into several low frequency components to extract the characteristics of electricity load more accurately and effectively. Later on, to compensate for the information loss during signal decomposition, the effect of exogenous external factors (i.e., weather variables) is incorporated in the forecasting model. To accomplish this task, we have introduced new correlation analysis technique i.e., T-Copula for: (i) determining the interdependence between electricity load and exogenous external factors, and (ii) deriving the peak load indicative threshold parameters from value at risk (VaR). The information from the signal decomposition and correlation analysis is employed to deep belief network (DBN) for final load forecasting.

4.3 Related Work of Hybrid STLF Models

If the load profile for a day is defined as $E_m(t) = [E_m(1), E_m(2), ..., E_m(N)]^T$, where $E_m(t)$ is the load profile on mth day and $t = 1, 2, 3, ..., N$ represents different time instances. The task of STLF model is to predict the load profile of future time instances i.e., $E_m(t+1)$ or $E_{m+1}(t)$. To avoid the notation complexity in later, we will use $E(t)$ as a load demand time series instead of load profile of a particular day $E_m(t)$.

In order to predict the energy load demand, researchers have proposed different hybrid models. For example, in [32], [33], different models such as back propagation neural network, genetic algorithm back propagation neural network, wavelet neural network, radical basis function neural network, general regression neural network, support vector machine are used separately to predict the energy load demand. Then, multi-objective flower pollination algorithm is applied to optimize the weight of each model. The final prediction value is determined from weighted average. Although the performance of the first category model is better than single model, there is a problem in calculating the weight of each model, which leaves a riddle for determining the optimal weights. Therefore, the second category model has been proposed by many researchers. For the second category model, electricity load is decomposed into several low frequency components. Then each component is predicted by a suitable model and the final forecasting result is the sum of each components forecasting results. Li et.al., [34], used wavelet transform to decompose the original electricity load into several components. Then each component is predicted by extreme learning machine combined with partial least squares regression. In [35], electricity load is decomposed by wavelet transform into some detailed sub series, and then each subseries is predicted by boundary network node model. Although wavelet transform can decompose the original electricity load into some low frequency components, it lacks the ability to extract the deep information as much as possible. To increase the efficiency of decomposition, in recent times EMD has been used by several researchers [36]–[40]. In [40] X.H. Qiu et.al., used EMD to decompose the electricity load into several intrinsic mode functions and one residual function. X.H. Qiu et.al., have predicted the future load demand and reported a higher load forecasting accuracy. However, the end effect and envelope fitting limitation associated with EMD decreases the efficiency of signal decomposition which consequently decreases the load forecasting accuracy. Besides, X.H. Qiu et.al., also ignored the effect of exogenous variables. Therefore, there is a scope to improve the load forecasting accuracy of [40]. In our proposed novel hybrid load forecasting model, our objective is to improve the load forecasting accuracy by: (i) suppressing the end effect and envelope fitting limitation of traditional EMD, and (ii) incorporating the effect of exogenous variables into the load forecasting model.

4.4 Framework of the Proposed Method

The framework of the proposed hybrid model for STLF is shown in Fig. 4.1. The basic architecture of the proposed hybrid model consist of load demand time series decomposition and processing of exogenous input variables with the help of correlation analysis. Load demand time series and exogenous input variables are processed in parallel. Compared to [40], application of IEMD will improve the signal decomposition efficiency and considering peak load indicative variable as input parameters will improve the load forecasting accuracy during peak load time. The binary peak indicative variable for each of the exogenous input is determined from the VaR computed from the correlation analysis of load demand with exogenous variables. This correlation analysis is done using T-Copula.

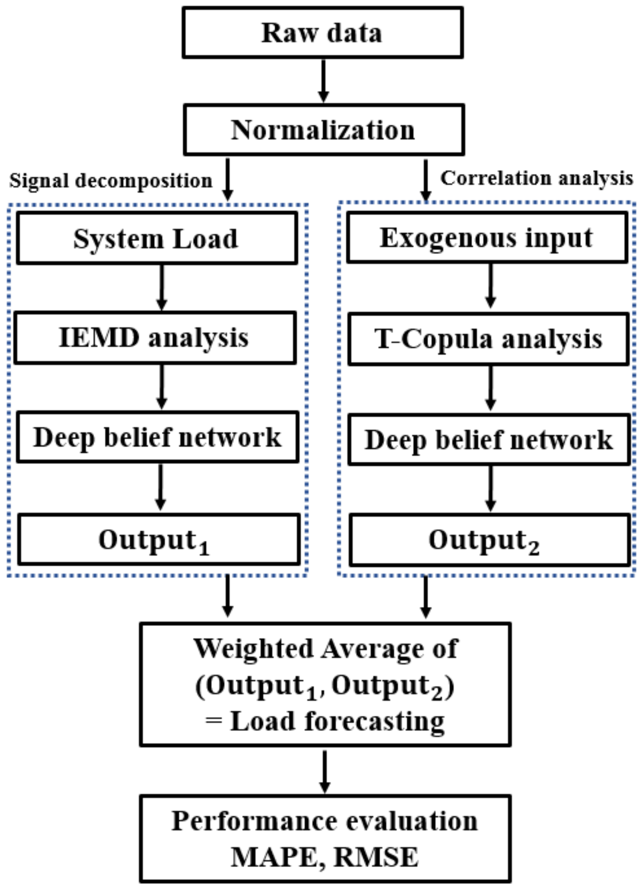

Figure 4.1. Framework of proposed hybrid STLF model.

The signal decomposition using IEMD will yield low frequency component called intrinsic mode functions (IMFs) e.g., $(IMF_1, IMF_2, IMF_3, ..., etc.,)$ and a signal monotone function i.e., residual function. The steps of the load forecasting from signal decomposition is given below:

- Step 1: In this step IEMD is employed to decompose the electricity load demand time series into different sub-series with different frequencies i.e., $(IMF_1, IMF_2, IMF_3, ..., etc.,)$, and a residual.

- Step 2: Each IMF and residual is forecasted using DBN, and the forecasting result of each of those is obtained.

- Step 3: The output obtained from each DBN are equally weighted and then aggregated to obtain $Output_1$.

When exogenous input variables are processed through T-Copula, the Gumbel-Hougaard Copula computes the upper tail dependence between energy load demand and the four exogenous input variables (e.g., dry bulb temperature, wet bulb temperature, dew point temperature, and humidity).

- Step 1: First, we will start with computing upper tail dependence correlation parameter $\lambda^u = [\lambda_1, \lambda_2, \lambda_3, \lambda_4]$ and the tandem parameters i.e., $VaR_1, VaR_2, VaR_3, VaR_4$ for each of the variables. Then the peak load indicative variable for each of the exogenous variables are determined from $VaR_1, VaR_2, VaR_3, VaR_4$.

- Step 2: Each of the DBN models are pre trained with the correlation parameter and peak load indicative variable. The forecasting result obtained for each of the exogenous variables.

- Step 3: The output obtained from each DBN are equally weighted and then aggregated to obtain $Output_2$.

4.5 Design Steps of the Proposed Method

4.5.1 Load Demand Time Series Signal Decomposition

There are several signal decomposition methods e.g., traditional wavelet transform, discrete wavelet transform, EMD. Compared to traditional wavelet transform EMD is highly preferable due to its applicability for non-stationary and nonlinear time series. However, there are some problems (such as end effect, envelope fitting) that needs to be controlled in EMD. IEMD is modification of EMD which is done by: (i) incorporating linear extrapolation to determine the end extremes so that the fitting envelope contain the given dataset, and (ii) employing nonuniform rational B-spline curve fitting envelope instead of cubic spline for processing complex signal. For clarification, first we presented the traditional EMD algorithm and it's issues as below [72]:

4.5.2 Traditional Empirical Mode Decomposition

EMD is an iterative shifting process which decomposes a signal into some regular low frequency components with different amplitude. The low frequency components include intrinsic mode functions (IMFs) and a residual function. The properties of the IMFs are given below:

(1) For each of the single IMF, the number of extrema and zero crossing throughout the whole length should be equal or differ by at most one.

(2) At any data location, the mean value of the envelope defined by local extrema is zero.

In order to satisfy those two properties, the iterative shifting process for extracting IMF from a given signal $E(t)$ is described below:

(1) Initially the local maxima $(E_{max}(t))$ and local minima $(E_{min}(t))$ of electricity load demand time series $E(t)$ are determined which are connected to construct upper and lower envelope with the help of cubic spline line.

(2) Then the difference between the mean of two envelopes and original load demand

time series is determined. If the average of the upper and lower envelope is denoted as $g_1(t)$, and the difference between $E(t)$ & $g_1(t)$ is defined as $d_1(t)$ then,

$$d_1(t) = E(t) - g_1(t) \tag{4.1}$$

In order to be an IMF, the $d_1(t)$ must obey the properties of IMF as mentioned above. Whenever $d_1(t)$ satisfies the conditions of IMF, then it is selected as first IMF $I_1(t)$. Else, the above steps are iteratively repeated.

(3) In the next step, the first IMF is subtracted from original electricity load demand time series to determine the residue $r_1(t)$,

$$r_1(t) = E(t) - I_1(t) \tag{4.2}$$

(4) Now the residue $r_1(t)$ is considered as new data subject to the shifting process as described above. Repeat the above process until the residue time series $r_1(t)$ is a monotone function i.e., residue data is small enough so that there is no turning point.

(5) By using the EMD, the original electricity load can be expressed as follows:

$$E(t) = \sum_{i=1}^{N} I_i(t) + r_n(t) \tag{4.3}$$

Following this iterative shifting process, the data can be represented by IMFs and a residual function.

4.5.3 Issue's With Traditional Empirical Mode Decomposition

Even though EMD algorithm decompose a complex time series more efficiently than other traditional decomposition techniques (e.g., wavelet transform or discrete wavelet transform), but EMD is associated with following issues [72], [73]:

(1) End effect of traditional EMD will cause divergent phenomena for both ends of the data. The end extremes of signal cannot be determined to be a maximum or a minimum. It makes the envelope distorted and affects the EMD decomposition. For example, once the first decomposed component is faulty, the latter decomposition will show the same results distortion. Thus, the obtained IMFs are not appropriate enough [73]. On the other hand, serious end effect will appear in the Hilbert transform of IMF which will form a spectral leakage. To enable the Hilbert spectrum and to reflect the characteristics of the original signals, we must suppress this issue effectively.

(2) Cubic spline fitting associated with traditional EMD will result in overshoot and undershoot phenomena. Thus the resulted envelope is not complete and consequently reflected into the extracted IMF.

4.5.4 Improved Empirical Mode Decomposition

To control the end effect and envelop fitting problem of traditional EMD, we proposed IEMD. By employing IEMD, we will suppress both end effect and envelope fitting limitations in the following way:

(1) Suppressing the end effect: In order to suppress the end effect and to achieve a real and effective decomposition, in this paper we have incorporated linear extrapolation to determine extreme ends of a signal so that the fitted envelope contain the given dataset. To

make a complete envelope which will contain the entire signal data, we must make a deal with the endpoint of the signal. More details can be found in [73].

The process by which this method determines the endpoints for upper envelope fitting is as shown in 4.2. Two maxima, A and B, are closest to an end. A straight line AB is linearly extended to the end point C. If point C is smaller than the endpoint value E of the signal, the point E is considered as a new maximum for the upper envelope fitting. Otherwise, if C is larger than the endpoint value E, point E is considered as a new maximum for the upper envelope intersection. Conversely we can determine the endpoints for lower envelope fitting.

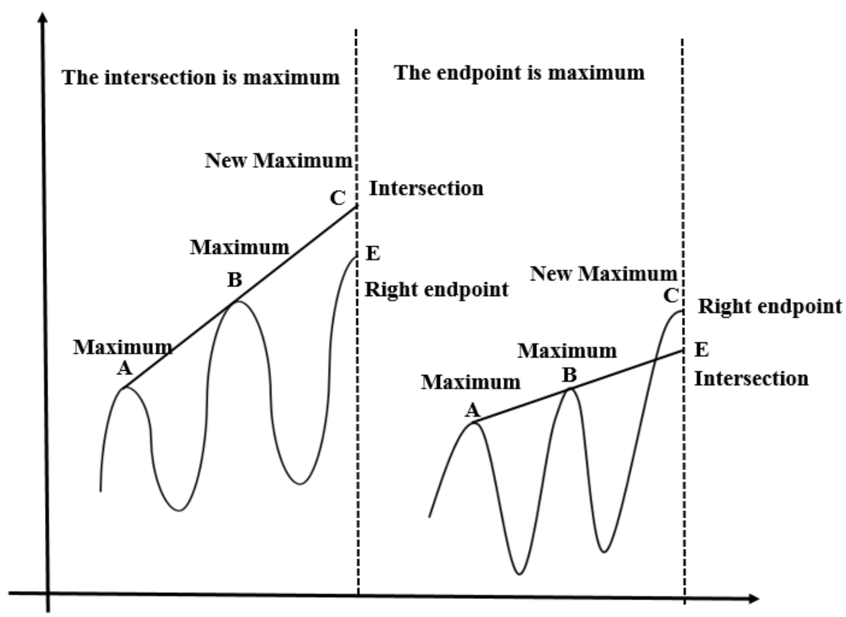

Figure 4.2. Determining the maxima of endpoint [73].

(2) Suppressing the envelope fitting: The original EMD algorithm proposed by Huang used cubic spline function to fit upper and lower envelope of the signal and then calculated the mean of the fitted upper & lower envelope. Because the power is low and easy to calculate, cubic spline curve fitting is simpler than others; however, the cubic spline fitting

will cause the overshoot and undershoot phenomena, so that the envelope fitting deviates from the actual signal envelope and develop a incomplete envelope. In order to solve the overshoot and undershoot problem of cubic spline curve fitting, many researchers has proposed improvement method, such as high order spline function method, polynomial fitting, and piecewise power function interpolation method. These methods can solve the problem of the fitting overshoot or fitting undershoot based on their own characteristics [73].

In this paper, a nonuniform rational B-spline fitting method is used to fit the upper and lower envelope of signal, resulting in the mean envelope. We use the accumulative chord length parameterized algorithm to achieve BNURBS curve fitting. The same simulation signal uses nonuniform rational B-spline (NURBS) curve fitting envelope compared with fitting envelope by cubic spline function. After employing IEMD, the simulation result of signal decomposition is shown in Fig. 4.3. With the decomposition results, it is obvious that IEMD algorithm can decompose the signal into different frequency components, and there is no mode mixing.

4.5.5 T-Copula Analysis

Preliminary research indicates that, there is upper tail dependence between power load and exogenous input variables. In this research, the Gumbel-Hougaard copula model computes the upper-tail dependence between the power load and the exogenous input variables (i.e., dry bulb temperature (D. bulb Temp.), wet bulb temperature (W. bulb Temp.), dew point temperature (D. point Temp.), and humidity). The classical bivariate Gumbel-Hougaard model can be defined as,

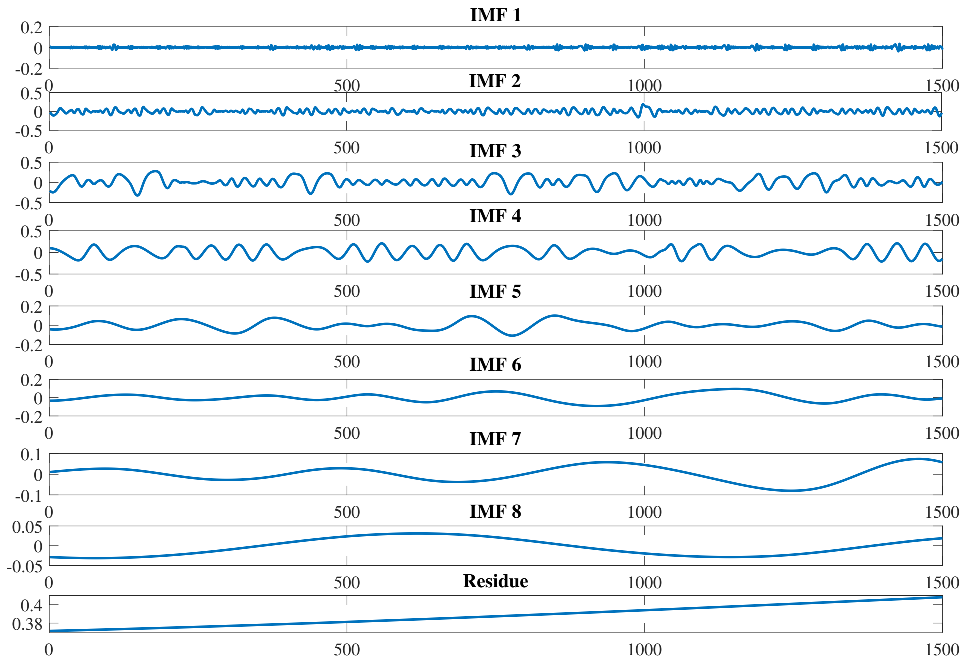

Figure 4.3. Signal decomposition using IEMD.

$$f(x_1(t), E(t)) = CP[f_{x_1}(x_1(t)), f_E(E(t))] \tag{4.4}$$

here $f_{x_1}(x_1(t))$ and $f_E(E(t))$ denotes the marginal cumulative distribution functions; x_1 represents one of the exogenous input variables, E denotes system load demand, $f(x_1,x_2)$ is the two dimensional joint distribution function; and $CP(x_1,E)$ is the Copula function. Now we need to determine the upper tail dependence parameter for each of the exogenous variables in the following way,

$$CP(x_1,E) = \exp\{-[(-\ln x_1)^{\alpha} + (-\ln E)^{\alpha}]^{1/\alpha}\} \tag{4.5}$$

The maximum likelihood method can be used to determine the copula model's parameter α. For the nonlinear relationship of system load demand and exogenous input variables, we adopt the Canonical Maximum Likelihood (CML) method that is implemented based on the empirical CDF of samples. The objective of the CML is expressed by:

$$\hat{\alpha} = \arg\min - \sum_{t=1}^{N} \ln f(x_1(t), E(t)) \tag{4.6}$$

here N indicates the number of exogenous input variables. Now upper-tail dependence parameter λ^1 of Gumbel-Hougard Copula is given by,

$$\lambda^1 = 2 - 2^{1/\alpha} \tag{4.7}$$

following this method we can determine our desired copula parameter for each of the exogenous input variables. Due to the variety of fluctuations and spikes of power load data, an effective statistical estimation of the peak load is crucial. In this research work, an indicative tandem variable called VaR is introduced to determine the peak load indicative variable for each of the variables. The computed peak indicative variables based on VaR helps to increase the load forecasting accuracy during peak load time. In our work, since exogenous input variables are stochastic and have impact on power load, we have determined the VaR from the following formula,

$$VaR_p^1 = CP^{-1}[f(x_1(t), E(t))] \tag{4.8}$$

here VaR_p^1 represents the pth upper percentile of bivariate distribution of exogenous input variable and system load. Hence, the two binary indicative of peak variable is determined from the following formula,

$$M(x_1) = \begin{cases} 1, & \text{if} x_1(t) \geq VaR_p^1 \\ \\ 0, & \text{if} x_1(t) < VaR_p^1 \end{cases} \tag{4.9}$$

here $M(x_1)$ represents the peak load indicative variable for one of the exogenous variables x_1 and the value of p is set as 0.95. For our research work, we will repeat this process for each of the exogenous input variables i.e., we need to do this calculation four times for four exogenous input variables.

In this research, the Gumbel-Hougaard Copula models fit the upper-tail dependence between system load versus exogenous weather variables (i.e., dry bulb temperature (D. bulb Temp.), wet bulb temperature (W. bulb Temp.), dew point temperature (D. point Temp.), and humidity). The default value of significance is set as 0.05 and the model parameters are estimated through maximum likelihood estimation. The upper-tail dependence parameter VaR between the system load and the dry bulb temperature is 3.12. For other parameters i.e., dew point temperature, wet bulb temperature, and humidity we got the VaR values of 1.69, 1.77, and 2.34 respectively. The calculation of VaR is done at 95th percentage of significance level. Strong upper-tail dependence of the power load on the exogenous weather variables have been shown in Fig. 4.4. To quantify the upper tail dependence parameter and to get idea about positive or negative correlation of system load with exogenous variables, in this paper we have added the Pearson correlation matrix as in Table. 4.1. This correlation matrix quantify the dependence and gives the idea about whether the exogenous variable has positive correlation or negative correlation.

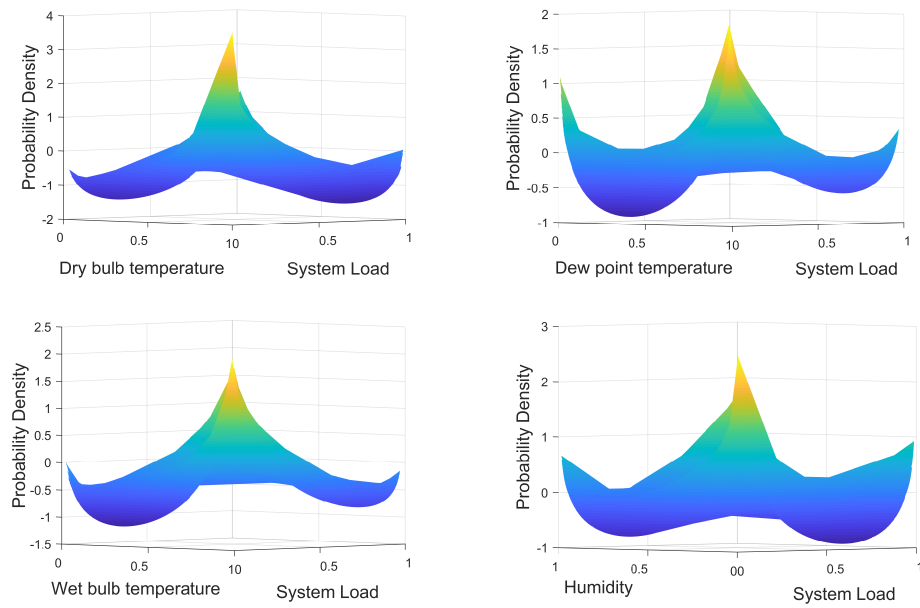

Figure 4.4. Correlation analysis of exogenous variables.

4.5.6 Learning with Deep Belief Network

A divide and conquer algorithm works by recursively breaking down a problem into two or more sub-problems of the same (or related) type, until these become simple enough to be solved directly. The solutions to the sub-problems are then combined to give a solution to the original problem. In the proposed method, the load demand data is decomposed into several IMFs and one residue by IEMD. Following the signal decomposition by IEMD, the data can be represented by IMF, to which we can employ Hilbert transform. The resulting Hilbert spectrum provides not only a more precise definition of particular events in time frequency space than wavelet transform but also more physically meaningful interpretation of the underlying dynamic processes. A DBN composed of a number of Restricted Boltzmann Machines (RBMs) and one ANN is applied to Hilbert spectrum of each IMF and the residue.

The DBN proposed by [74] provides a new way to train deep generative models, which

Table 4.1. Pearson Correlation Matrix for correlation analysis between system load and input exogenous variables

Pearson Correlation matrix.					
	D. bulb Temp.	D. point Temp.	W. bulb Temp.	Hum.	Sys. Load
D. bulb Temp.	1.00	0.64	0.89	-0.25	0.10
D. point Temp.	0.64	1.00	0.91	0.56	-0.11
W. bulb Temp.	0.89	0.91	1.00	0.20	-0.02
Hum.	-0.25	0.56	0.20	1.00	-0.27
Sys. Load	0.10	-0.11	-0.02	-0.27	1.00

is called layer-wise greedy pre-training algorithm. Fig. 4.5 shows the architecture of a DBN. There is no inter-connection between units in each layer. A RBM is a neural network which can learn the probability distribution over the input dataset. The DBN pre-training procedure treats each consecutive pair of layers in the multi layer perceptron (MLP) as a RBM [75] whose joint probability is defined as,

$$P_{h|v}(h|v) = \frac{1}{Z_{h,v}} * e^{(v^T Wh + v^T b + a^T h)} \tag{4.10}$$

Here, h represents input applied to hidden layer, v represents output obtained from visible layer, W represents hidden neuron weights, and a represents activation. For each RBM there is pair of hidden layer and visible layer. For the Bernoulli–Bernoulli RBM applied to binary v with a second bias vector b and normalization term $Z_{h,v}$, and

$$P_{h|v}(h|v) = \frac{1}{Z_{h,v}} * e^{(v^T Wh + (v-b)^T (v-b) + a^T h)} \tag{4.11}$$

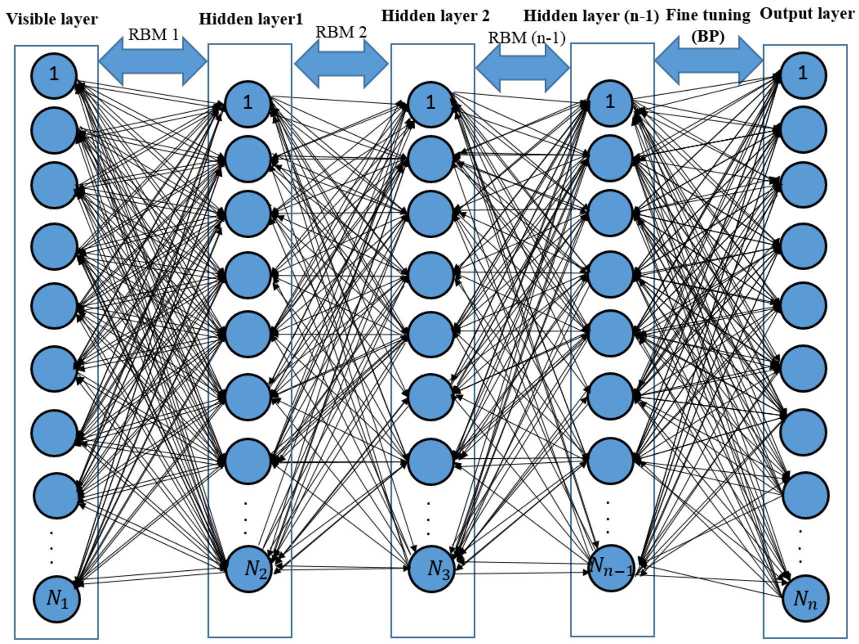

Figure 4.5. Deep belief network architecture.

for the Gaussian–Bernoulli RBM applied to continuous variable v [76]. In both cases the conditional probability $P_{h|v}(h|v)$ has the same form as that in an MLP layer.

The objective function of an RBM is,

$$L(a,b,W) = \Sigma \log P_{h|v}(h|v) \tag{4.12}$$

The layer-wise pre-training method requires the DBN to be pre-trained following a stochastic gradient descent method on the objective function. The gradient method indicates that, the parameters $(e.g., a, b, W)$ are updated based on the gradients of the objective function 4.12. The gradients of the probability distribution function can be expressed in the following way,

$$\frac{\partial P_{h|v}(h|v)}{\partial W_{j,i}} = \langle v_i h_i \rangle_{P_{h|v}(h|v)} - \langle h_i v_i \rangle_{recon} \tag{4.13}$$

$$\frac{\partial P_{h|v}(h|v)}{\partial a_i} = \langle v_i \rangle_{P_{h|v}(h|v)} - \langle v_i \rangle_{recon} \tag{4.14}$$

$$\frac{\partial P_{h|v}(h|v)}{\partial b_i} = \langle h_i \rangle_{P_{h|v}(h|v)} - \langle h_i \rangle_{recon} \tag{4.15}$$

here $\langle h_i \rangle_{P_{h|v}(h|v)}$ is the expectation of the conditional distribution with respect to the input raw data; $\langle h_i v_i \rangle_{recon}$ is the expectation of the i^{th}-step reconstructed distribution. We can use contrastive divergence [77] to obtain the expectation of the reconstructed distribution through alternating Gibbs sampling. Later, we used the following updating formulas,

$$W_{i+1} = W_i + \eta(\langle v_i h_i \rangle_{P_{h|v}(h|v)} - \langle h_i v_i \rangle_{recon}) \tag{4.16}$$

$$a_{i+1} = a_i + \eta(\langle v_i \rangle_{P_{h|v}(h|v)} - \langle v_i \rangle_{recon}) \tag{4.17}$$

$$b_{i+1} = b_i + \eta(\langle h_i \rangle_{P_{h|v}(h|v)} - \langle h_i \rangle_{recon}) \tag{4.18}$$

To train multiple layers, one trains the first layer, freezes it, and uses the conditional expectation of the output as the input to the next layer and continues training next layers. Based on the layer wise pre-training approach, all the parameters of the DBN algorithm are initialized. Hinton and many others have found that initializing MLPs with pretrained parameters never hurts and often helps [74], [78]. Adjustment of these parameters in a supervised manner is conducted until the loss function of the DBN reaches its minimum.

Finally, back-propagation algorithm is applied for the fine-tuning process. All parameters are updated from the top to bottom resulting reduced forecasting errors.

Due to the influence from climate and social activities, the electricity load data shows three main nest cycles: daily, weekly and yearly. To identify cycles and patterns in load demand time series data, autocorrelation function (ACF) can be applied as a guidance for informative feature subset selection [79]. Suppose a time series data set is given as $E = E_t : t \in T$, where T is the index set. The lag k autocorrelation coefficient r_k can be computed by:

$$r_k = \frac{\sum_{t=k+1}^{n}(E_t - \overline{E})(E_{t-k} - \overline{E})}{\sum_{t=1}^{n}(E_t - \overline{E})^2} \tag{4.19}$$

where $\overline{E}$ is the mean value of all E in the given time series, r_k measures the linear correlation of the time series at times t and k.

4.6 Simulation Results and Analysis

4.6.1 Description of the Dataset

The proposed hybrid load forecasting model is validated on the Australian Energy Market Operator (AEMO) data [69] and the dataset for one of urbanized regions of Houston, Texas, USA [70]. Specifically, the dataset include three main groups of measured variables: weather data (i.e., dry bulb temperature, wet bulb temperature, dew point temperature, and humidity), time categorical data (i.e., hour, month, day), social data (i.e., working day, weekend, holiday), and energy load demand for specific sampling time.

4.6.2 Performance Evaluation Criteria

The performance of the proposed load forecasting models are compared with respect to mean absolute percentage error (MAPE) and root mean square error (RMSE) [23], [72].

1) MAPE is defined as,

$$MAPE = (1/N) * \sum_{t=1}^{N} \frac{\mid E(t) - \hat{E}(t) \mid}{\mid E(t) \mid} * 100 \tag{4.20}$$

here $E(t)$ denotes actual load demand and $\hat{E}(t)$ denotes the forecasted load demand.

2) RMSE is defined as,

$$RMSE = \sqrt{\frac{\sum_{t=1}^{N} \mid E(t) - \hat{E}(t) \mid^2}{N}} \tag{4.21}$$

The values of MAPE and RMSE provides the idea about forecasting accuracy. The less the values of MAPE and RMSE means higher forecasting accuracy.

4.6.3 Experimental Results

All the simulations are conducted using Matlab R2017b on a standard PC. The results are validated for two case study result. The dataset for two case study are: (i) AEMO data, Autsralia and (ii) Dataset of Houston, Texas, USA. For both of the case study, we have normalized the dataset scaled into $[0, 1]$ using the following formula,

$$\bar{E}_i = \frac{E_{max} - E_i}{E_{max} - E_{min}} \tag{4.22}$$

(1) Case Study #1: In this case study we have collected dataset from AEMO [69]. The data collection date is from 1st January 2013 to 31st December 2013 with sampling time of

half hour. We have divided the whole year dataset into four seasons: (i) January to March, (ii) April-June, (iii) July-September, and (iv) October-December. During the training time we have considered the immediate last three days load demand as a batch for forecasting the next day load demand. At the same time we have considered the weather information of that day for load demand forecasting. Following this method for week ahead load forecasting we have considered the three week dataset of a month as training dataset and remaining week as the testing dataset. But here we assumed that we have the information of day-type i.e., the working day or holiday, time of day for avoiding uncertainty due to volatile nature of electricity load and exogenous input variables. The input dataset obtained from signal decomposition is a frequency spectrum. These frequency spectrum sub series data are subjected to Hilbert Huang frequency spectrum transform and then applied to DBN for training and testing. The input dataset obtained from correlation analysis includes upper tail dependence parameter, binary peak indicative variable and lag autocorrelation parameter. These variables are applied to DBN for load demand prediction due to exogenous variable. For a fair comparison, each month of 2013 is considered as training and testing dataset. The prediction result of the proposed model is compared with [72]. For week ahead load prediction, we have taken three weeks of a month as training dataset and remaining week dataset as testing dataset. Notice that, here we have trained the DBN with batch of three days data of same time and similar day i.e., working day or weekend.

Therefore we have load demand prediction from signal decomposition and correlation analysis. For making final prediction i.e., forecasted load demand we have aggregated the results from signal decomposition and correlation analysis. We have considered the equal weighted average to determine the final forecasted load demand. The forecasted load de-

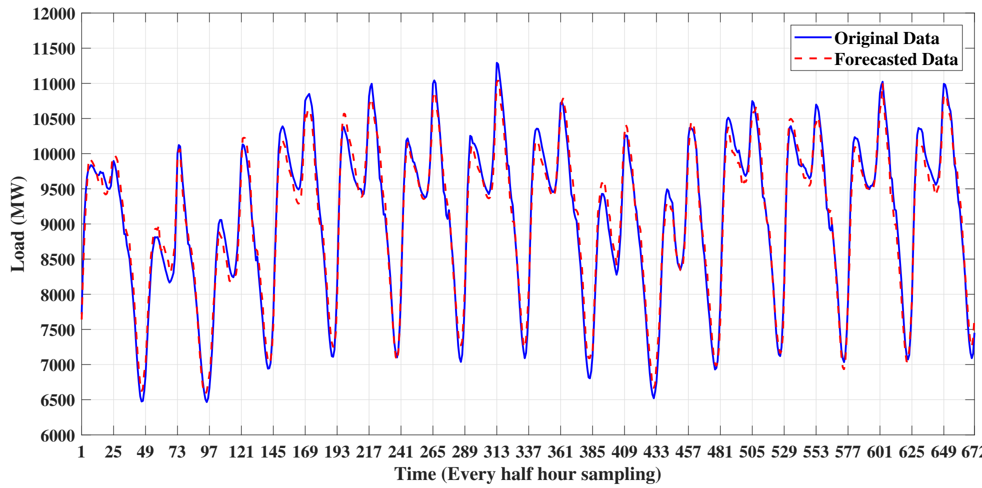

Figure 4.6. Two Week-Ahead Load Forecasting Result.

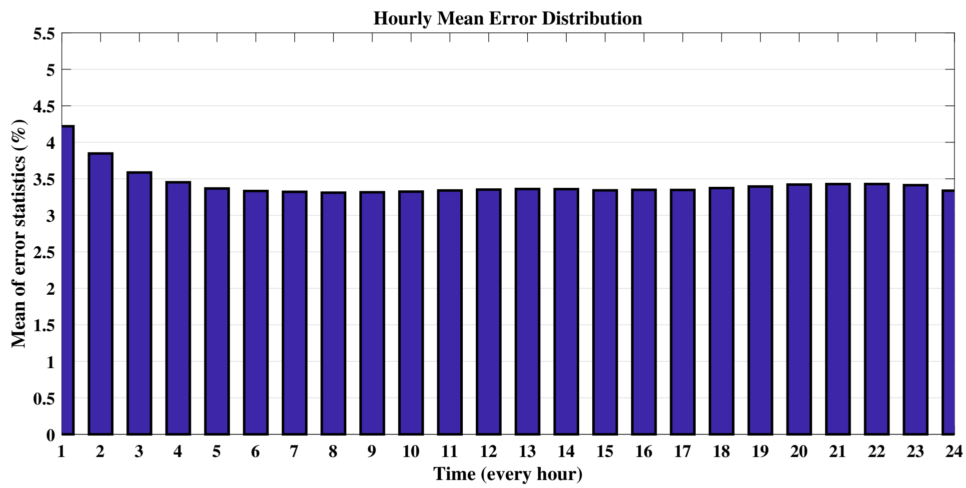

Figure 4.7. Hourly Error Distribution.

mand from the proposed model is shown in Fig. 4.6. The simulation result presented in Fig. 4.6 is carried out in New South Wales (NSW), Australia during the month January- March 2013. Dataset of year 2013 is considered for comparison with [40]. We have presented the mean error distribution at every hour as shown in Fig. 4.7. This error distribution is presented to show the load forecasting accuracy improvement during peak load time. From the mean error distribution result it is evident that, there is a improvement in load forecasting accuracy during peak time and this will help the utility operators to make proper

Table 4.2. Load Forecasting Performance Comparison: Case Study 1, Location five regions of Australia.

Month	Algorithm	NSW		TAS		QLD		VIC		SA	
		MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE
Jan.-March	NN [40]	6.16	587.82	6.39	91.56	4.85	409.51	8.56	759.38	12.97	225.75
	DBN [40]	6.05	633.11	6.18	86.24	4.53	348.71	6.26	465.28	11.02	202.64
	EMD-DBN [40]	4.62	541.53	4.05	56.10	2.56	191.22	8.86	762.57	10.04	238.09
	Coupla-IEMD-DBN	**3.44**	**392.44**	**2.98**	**39.81**	**2.28**	**186.90**	**6.96**	**598.76**	**7.58**	**181.91**
April-June	NN [40]	6.64	743.44	8.27	123.11	5.65	399.51	9.06	569.34	13.76	201.87
	DBN [40]	6.44	699.17	7.32	102.71	5.27	369.54	6.44	426.41	11.21	182.81
	EMD-DBN [40]	3.22	377.63	5.80	85.13	2.93	243.68	4.35	321.59	6.76	125.31
	Coupla-IEMD-DBN	**2.41**	**314.03**	**4.34**	**66.42**	**2.39**	**201.32**	**3.69**	**276.51**	**5.14**	**98.72**
July-Sept.	NN [40]	7.64	732.24	8.70	161.28	5.38	372.26	8.39	728.06	14.19	381.68
	DBN [40]	5.17	480.69	6.48	119.53	5.11	357.87	7.85	546.61	11.43	223.44
	EMD-DBN [40]	3.08	322.04	4.93	73.91	2.08	142.84	3.83	285.45	9.60	192.74
	Coupla-IEMD-DBN	**2.42**	**271.21**	**3.84**	**58.32**	**1.68**	**118.36**	**3.41**	**228.89**	**7.54**	**158.08**
Oct.-Dec.	NN [40]	7.88	796.73	6.89	165.8	5.44	374.92	7.27	520.11	13.86	391.23
	DBN [40]	6.62	785.3	5.96	95.41	5.53	388.71	6.88	561.05	11.66	386.82
	EMD-DBN [40]	2.71	282.34	4.75	68.26	2.88	219.19	3.73	322.91	8.11	192.74
	Coupla-IEMD-DBN	**2.18**	**224.52**	**3.78**	**56.56**	**198.56**	**3.12**	**3.12**	**291.35**	**5.86**	**154.32**

generation scheduling and distribution maintenance planning. And for comparison with results presented in [40], we have done simulation for all of the regions of Australia as given in Table. 4.2. As seen in Table. 4.2, error in load forecasting results i.e., MAPE and RMSE values of the proposed model are lower than the other comparative models in [40]. The MAPE values of the proposed model are decreased by 21.19%, and the RMSE values decreased by 16.93% compared to [40]. The reason of performance improvement is due to : (i) IEMD signal decomposition, and (ii) T-Copula correlation analysis. IEMD improves the signal decomposition efficiency and T-Copula contributes to improve the load forecasting accuracy during peak time by computing peak load indicative variables from VaR.

(2) Case Study #2: For this case study we have collected the dataset from urbanized area of Houston, Texas, USA [70]. The data collection date is from 1st January 2016 to 31st December 2016 with sampling time of one hour. We have divided the whole year dataset into four seasons: (i) January to March, (ii) April-June, (iii) July-September, and

(iv) October-December. During the training time we have considered the last immediate three days dataset as a batch for forecasting the load demand of next day. Following the similar procedure as mentioned for case study #1, for this case study we have again considered the three week dataset of a month as training dataset and remaining week as the testing dataset. The input dataset obtained from signal decomposition is a frequency spectrum. These frequency spectrum sub series data are subjected to Hilbert Huang frequency spectrum transform and then applied to DBN for training and testing. The input dataset obtained from correlation analysis includes upper tail dependence parameter, binary peak indicative variable and lag autocorrelation parameter. These variables are applied to DBN for load demand prediction due to exogenous variable.

Table 4.3. Load Forecasting Performance Comparison: Case Study 2, Location Houston, Texas, USA.

		Location: Houston	
Month	Algorithm	MAPE	RMSE
Jan.-March	NN [[80]]	7.37	2521.19
	DBN [[80]]	6.99	2483.34
	Copula-DBN [[80]]	6.08	2263.61
	Coupla-IEMD-DBN	**4.11**	**2014.18**
April-June	NN [[80]]	8.16	1593.68
	DBN [[80]]	7.78	1479.97
	Copula-DBN [[80]]	6.63	1388.84
	Coupla-IEMD-DBN	**4.62**	**1325.30**
July-Sept.	NN [[80]]	7.19	2521.19
	DBN [[80]]	6.88	2230.02
	Copula-DBN [[80]]	6.21	2017.42
	Coupla-IEMD-DBN	**3.98**	**1940.68**
Oct.-Dec.	NN [[80]]	8.25	2213.67
	DBN [[80]]	7.99	2203.74
	Copula-DBN [[80]]	7.15	2110.45
	Coupla-IEMD-DBN	**5.46**	**1856.86**

The prediction result of the proposed model is compared with [80] and the results are presented in Table. 4.3. As seen in Table. 4.3, all MAPE and RMSE values of the proposed model are lower than the traditional EMD based STLF model. The MAPE values of the proposed model are decreased by 15.27%, and the RMSE values decreased by 13.86% compared to [80]. This significant decrease in MAPE and RMSE values resulted from the combined effect of IEMD and T-Copula. These two method enables our proposed hybrid model for processing more information.

4.7 Summary

This chapter presented a novel hybrid STLF model. First, load demand time series is decomposed by IEMD. Second, correlation analysis between system load and exogenous input variables are incorporated to increase the load forecasting accuracy during peak time. Third, the two components are predicted separately by the suitable model. Last, each component's forecasting results are added up to obtain the final forecasting results. Electricity load data from Australia and Texas electricity markets are used to validate the effectiveness of the proposed model. All case study results indicate that the proposed model improves the forecasting accuracy. Three facts emerge clearly from the results: (1) the linear and nonlinear component of electricity load can be extracted more accurately and effectively by the IEMD, (2) the peak load indicative variable computed from VaR through T-Copula model improves the load forecasting accuracy during peak time, (3) the DBN has a strong ability to fit the nonlinear component of the original electricity load. By using each model's advantage, the hybrid model can capture the different characteristics associated with electricity load. Therefore, the proposed model can provide a robust, stable and more accurate prediction results.

CHAPTER 5 CONCLUSIONS AND FUTURE WORK

5.1 Conclusions and Discussions

Nowadays distributed energy sources are considered as viable solution for meeting the cumulative increase of energy load demand. The task of load profile data analytics is extremely important for energy demand management, stability and security of power systems. However, increasing penetration of intermittent and variable renewable energy sources has significantly complicated data analytics. The load profile data analytics deals with load profile classification, bad data identification, STLF etc. A sufficiently accurate, robust and fast STLF model is necessary for the day-to-day reliable operation of the grid. This thesis has presented some load profile data analytics and data driven STLF models.

First, the load profile of customers are volatile and correlated with several factors e.g., weather variables, social variables. However, load profile for different customers shows some similarities with in the same time frame. In order to make operating decisions in modern day automated power system operation, the utility operators needs to group the similar consumer's into same group i.e., consumer load profile classification is needed. In the previous study, authors have provided different kNN based load profile classification. In our proposed work, we suppressed the influence of irrelevant feature of kNN based models and provided an insight to predict the load demand variability of a consumer. In our work, we have introduced recently developed ENN model for load profile classification and computed the generalized class-wise statistics to predict the load demand variability of individual customers. In addition, we have provided an insight to estimate the load demand. The classification accuracy is improved significantly with respect to kNN based load profile classification.

Second, the non-stationarity, non-linearity, and multiple-seasonality characteristics of load demand time series makes the STLF a complicated task. This difficulty is conventionally tackled with data-driven methodologies that require domain-specific knowledge. However, the ideal choice of a data-driven methodology that extracts relevant and meaningful information from available data even when the physical model of the system is unknown. Our work is focused on developing a data-driven composite ENN model for STLF. The proposed composite model efficiently identifies the characteristics of load data that are affected by multiple exogenous factors including time, day, weather, seasons, social activities, and economic aspects. The effectiveness of the proposed method is evaluated and observed to be competing with the benchmark methods. The satisfactory performance suggests that the proposed data-driven model can be used as an effective tool for the real-time STLF task.

Third, in the context of non-stationarity, non-linearity, and multiple-seasonality of load demand time series a single model is inadequate to represent the inherent characteristics of load demand time series. Therefore, different hybrid models have been proposed to represent the inherent characteristics of load demand time series. However, combining different models for making final prediction requires rigorous calculations to determine the weights of different models. To this end, we have introduced load demand time series signal decomposition and correlation analysis between system load with different exogenous input variables. In our proposed hybrid STLF models, we have suppressed the end effect and envelope fitting limitation of traditional EMD. We have also computed the peak load indicative variables to improve the load forecasting accuracy during peak time. The accuracy of hybrid STLF models is significantly improved.

Overall, we have presented the load profile data analytics for facilitating the load monitoring control unit. The idea of load demand variability is the key information for load monitoring control unit. The proposed load forecasting models will help to home energy management research. Based on more accurate forecasted load demand we can develop different optimization techniques for demand response applications.

5.2 Future Work

The future work along this direction includes the following major tasks:

1. Future work will focus on including customer related information and external factors for load profile classification. The research aim will be to find the customers behaviour mode behind load features. This research can then be used for accurate demand forecasting, supporting efficient use of demand response program, and enhancing supplier settlement efficiency.

2. Future research will focus on integrating more customer information and external factors for load forecasting to find the customers' behavior mode behind load features. Potential applications of this load forecasting includes generation units scheduling which helps to optimize demand side management system.

3. Advanced models can be used to select suitable input variables for electricity load forecasting in the future. Besides, some other future influencing factors such as information of consumer related to incentive based demand response program, and uncertainty from distributed renewable energy integration can be added in the hybrid

model as future research. This framework can be beneficial for practical short-term generation scheduling and operations for the grid network..

In general, all these works are expected to enhance the power system quality, stability and reliability by improving the load forecasting accuracy.

LITERATURE CITED

[1] Z. Ni and A. Das, "A new incentive-based optimization scheme for residential community with financial trade-offs," *IEEE Access*, vol. 6, pp. 57 802–57 813, 2018, ISSN: 2169-3536. DOI: 10.1109/ACCESS.2018.2872929.

[2] Z. Ni and A. Das, "A new incentive-based optimization scheme for residential community with financial trade-offs," *IEEE Access*, vol. 6, pp. 57 802–57 813, 2018, ISSN: 2169-3536. DOI: 10.1109/ACCESS.2018.2872929.

[3] R. Hu, S. Wen, Z. Zeng, and T. Huang, "A short-term power load forecasting model based on the generalized regression neural network with decreasing step fruit fly optimization algorithm," *Neurocomputing*, vol. 221, pp. 24 –31, 2017, ISSN: 0925-2312. DOI: https://doi.org/10.1016/j.neucom.2016.09.027. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S092523121631044X.

[4] T. Dang-Ha, F. M. Bianchi, and R. Olsson, "Local short term electricity load forecasting: Automatic approaches," in *2017 International Joint Conference on Neural Networks (IJCNN)*, 2017, pp. 4267–4274. DOI: 10.1109/IJCNN.2017.7966396.

[5] T. Roy, A. Das, and Z. Ni, "Optimization in load scheduling of a residential community using dynamic pricing," in *2017 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, 2017, pp. 1–5. DOI: 10.1109/ISGT.2017.8086087.

[6] S. Fan and R. J. Hyndman, "Short-term load forecasting based on a semi-parametric additive model," *IEEE Transactions on Power Systems*, vol. 27, no. 1, pp. 134–141, 2012, ISSN: 0885-8950. DOI: 10.1109/TPWRS.2011.2162082.

[7] T. Hong and S. Fan, "Probabilistic electric load forecasting: A tutorial review," *International Journal of Forecasting*, vol. 32, no. 3, pp. 914–938, 2016, ISSN: 0169-2070. DOI: "https://doi.org/10.1016/j.ijforecast.2015.11.011". [Online]. Available: "http://www.sciencedirect.com/science/article/pii/S0169207015001508".

[8] T. Dang-Ha, F. M. Bianchi, and R. Olsson, "Local short term electricity load forecasting: Automatic approaches," in *2017 International Joint Conference on Neural Networks (IJCNN)*, 2017, pp. 4267–4274. DOI: 10.1109/IJCNN.2017.7966396.

[9] A. Al-Kandari, S. Solaiman, and M. El˙Hawary, "Fuzzy short-term electric load forecastingl," *International Journal of Electrical Power & Energy Systems*, vol. 26, no. 2, pp. 111–122, 2004.

[10] Y. He, R. Liu, H. Li, S. Wnag, and X. Lu, "Short-term power load probability density forecasting method using kernel-based support vector quantile regression and copula theory," *Applied Energy*, vol. 185, no. 1, pp. 254–266, 2017.

[11] C. -N. Ko and C. Lee, "Short-term load forecasting using svr (support vector regression)-based radial basis function neural network with dual extended kalman filter," *Energy*, vol. 49, no. 1, pp. 413–422, 2013.

[12] A. H. A. H. J. Sadaei R. Enayatifar and A. Gani, "Short-term load forecasting using a hybrid model with a refined exponentially weighted fuzzy time series and an improved harmony search," *International Journal of Electrical Power& Energy Systems*, vol. 62, pp. 118–129, 2014.

[13] J. M. A. B. C. A. S.-E. L. Hernández C. Baladrón and J. Lloret, "Artificial neural networks for short-term load forecasting in microgrids environment," *Energy*, vol. 75, pp. 252–264, 2014.

[14] Q. X. Y. C. P. Li Y. Li and Y. Zhang, "Application of a hybrid quantized elman neural network in short-term load forecasting," *International Journal of Electrical Power & Energy Systems*, vol. 55, pp. 749–759, 2014.

[15] A. K. M. Rana I. Koprinska and V. Agelidis, "Prediction intervals for electricity load forecasting using neural networks," *2013 International Joint Conference on Neural Network (IJCNN)*, pp. 1–8, 2013.

[16] G. Sudheer and A. Suseelatha, "Short term load forecasting using wavelet transform combined with holt–winters and weighted nearest neighbor models," *International Journal of Electrical Power & Energy Systems*, vol. 64, pp. 340–346, 2015.

[17] S.-J. Huang and K.-R. Shih, "Short-term load forecasting via arma model identification including non-gaussian process considerations," *IEEE Transactions on Power Systems*, vol. 18, pp. 673–679, 2003.

[18] C.-M. Lee and C.-N. Ko, "Short-term load forecasting using lifting scheme and arima models," *Expert Systems with Applications*, vol. 38, no. 5, pp. 5902 –5911, 2011.

[19] H. S. A. Kavousi-Fard and F. Marzbani, "A new hybrid modified firefly algorithm and support vector regression model for accurate short term load forecasting," *Expert Systems with Applications*, vol. 41, no. 13, pp. 6047 –6056, 2014.

[20] A. A. A. V. A. Khwaja M. Naeem and B. Venkatesh, "Improved short-term load forecasting using bagged neural networks," *Electric Power Systems Research*, vol. 125, pp. 109–115, 2015.

[21] P. R. P. S. C. Cecati J. Kolbusz and B. M. Wilamowski, "A novel rbf training algorithm for short-term electric load forecasting and comparative studies," *IEEE Transactions on Industrial Electronics*, vol. 62, pp. 6519–6529, 2015.

[22] Y. Z. L. Wang and T. Chen, "Back propagation neural network with adaptive differential evolution algorithm for time series forecasting," *Expert Systems with Applications*, vol. 42, no. 2, pp. 855 –863, 2015.

[23] R. Zhang, "Short-term load forecasting of australian national electricity market by an ensemble model of extreme learning machine," *IET Generation, Transmission & Distribution*, vol. 7, no. 6, pp. 391–397, 2013.

[24] F. Z. G.M. Khan A.R. Khattak and S. Mahmud, “Electrical load forecasting using fast learning recurrent neural networks,” *2013 International Joint Conference on Neural Network (IJCNN)*, pp. 1–6, 2013.

[25] S. R. M. P. F. Fahiman S.M. Erfani and C. Leckie, “Improving load forecasting based on deep learning and k-shape clustering,” 2016, pp. 1–5.

[26] Z. Y. D. W. K. R. Zhang Y. Xu and K. P. Wong, “A composite k-nearest neighbor model for day-ahead load forecasting with limited temperature forecasts,” 2017, pp. 4134–4141.

[27] C.-J. H. C.-M. Huang and M.-L. Wang, “A particle swarm optimization to identifying the armax model for short-term load forecasting,” vol. 18, 2003, pp. 673–679.

[28] D. H. Kyung. SB Young. SB and G. J., “Short term load forecasting for the holidays using fuzzy linear regression method,” *IEEE Transactions on Power Systems*, vol. 1, no. 1, pp. 96 –101, 2005.

[29] C. JX and W. JZ, “Short-term load forecasting using a kernel based support vector regression combination model,” *Applied Energy*, vol. 132, pp. 602 –609, 2014.

[30] L. F. Jiang. P and S. YL, “A hybrid forecasting model based on date-framework strategy and improved feature selection technology for short-term load forecasting,” *Energy*, vol. 119, pp. 694 –709, 2017.

[31] C. F and K. M., “A seasonal direct optimal hybrid model of computational intelligence and soft computing techniques for electricity load forecasting,” *Energy*, vol. 140, pp. 988 –1004, 2017.

[32] Y. M. M. J. Xiao. LY Shao. W and J. CJ., “Research and application of a combined model based on multi-objective optimization for electrical load forecasting,” *Energy*, vol. 119, pp. 1057 –1074, 2017.

[33] W. C. Z. K. Xiao. LY Shao. W and L. Hy., “Research and application of a hybrid model based on multi-objective optimization for electrical load forecasting,” *Energy*, vol. 180, pp. 213 –233, 2017.

[34] G. L. Li. S and W. P., “An ensemble approach for short-term load forecasting by extreme learbing machine,” *Applied Energy*, vol. 170, pp. 22 –39, 2016.

[35] A. A. Ghofrani. M Ghayekhloo. M and G. A., “A hybrid short term load forecasting with new input selection framework,” *Energy*, vol. 170, pp. 777 –786, 2017.

[36] D. H. Kyung. SB Young. SB and G. J., “Short term load forecasting for the holidays using fuzzy linear regression method,” *IEEE Transactions on Power Systems*, vol. 1, no. 1, pp. 96 –101, 2005.

[37] W. H. H. W. Fan. GF Qing. S and L. HJ., “Support vector regression model based on empirical mode decompsition and auto regression for electric load forecasting,” *Energies*, vol. 6, no. 1, pp. 1887 –1901, 2016.

[38] B. Y. Xiong. T and H. Z., "Interval forecasting of electricity demand: A novel bivariate emd-based support vector regression modeling framework," *International Journal of Electric Power Energy System*, vol. 63, pp. 353 –362, 2014.

[39] Z. J. F. W. Liu. N Tang. QF and L. J., "A hybrid forecasting model with parameter optimization for short term load forecasting of microgrids," *Applied Energies*, vol. 129, pp. 336 –345, 2014.

[40] P. S. X.H. Qiu Y. Ren and G. Amaratunga., "Empirical mode decomposition based ensemble deep learning framework for load demand time series forecasting," *Applied Soft Computing*, vol. 54, no. 6, pp. 246 –255, 2017.

[41] Q. W. Z. H. J. H. K. Chen K. Chen and J. He, "Short-term load forecasting with deep residual networks," *IEEE Transactions on Smart Grid*, Early Access.

[42] D. Bunn and E. Farmer, "Comparative models for electric load forecasting," *New York, NY, USA: Wiley*, vol. 1st ed, 1985.

[43] D. D. G. Cerne and I. Skrjanc, "Short-term load forecasting by separating daily profiles and using a single fuzzy model across the entire domain," *IEEE Transactions on Industrial Electronics*, vol. 65, pp. 7406–7415, 2018.

[44] S. R. A. Mutanen M. Ruska and P. Jarventausta, "Customer classification and load profiling method for distribution systems," *IEEE Transactions on Power Delivery*, vol. 26, pp. 1755–2151, 2011.

[45] B. Tang and H. He, "2016 ieee power and energy society general meeting (pesgm)," *IEEE Computational Intelligence Magazine*, vol. 10, pp. 52–60, 2015.

[46] M. S.-E.-E. N. Mahmoudi-Kohan M. P. Moghaddam and E. Shayesteh, "A three-stage strategy for optimal price offering by a retailer based on clustering techniques," *International Journal of Electrical Power & Energy Systems*, vol. 32, no. 10, pp. 1135 –1142, 2010.

[47] X. W. P. Zhang X. Wu and S. Bi, "Short-term load forecasting based on big data technologies," *CSEE Journal of Power and Energy Systems*, vol. 1, pp. 59–67, Sept. 2015.

[48] M. K. S.-E.-E. N. Mahmoudi-Kohan M. P. Moghaddam and S. M. Bidaki, "Improving wfa k-means technique for demand response programs applications," *2009 IEEE Power Energy Society General Meeting*, pp. 1 –5, 2009.

[49] I. M. J. I. G.-J. B. C. Leon F. Biscarri and R. Millan, "Variability and trend-based generalized rule induction model to ntl detection in power companies," *IEEE Transactions on Power Systems*, vol. 26, pp. 1798–1807, 2011.

[50] W. G. Z. D.-h. Liu Li, "Application of k-means clustering algorithm in load curve classification," *Power system protection and control*, vol. 26, no. 4, pp. 65–68, 2011.

[51] S. l. Y. K. le Zhou and C. Shen, “A review of electric load classification in smart grid environment,” *Renewable and Sustainable Energy Reviews*, vol. 24, pp. 103–110, 2013.

[52] C. K. Y. Wang Q. Chen and Q. Xia, “Clustering of electricity consumption behavior dynamics toward big data applications,” *IEEE transactions on smart grid*, vol. 7, no. 5, pp. 2437–2447, 2016.

[53] F. L. R. Li and N. D. Smith, “Multi-resolution load profile clustering for smart metering data,” *IEEE Transactions on Power Systems*, vol. 31, pp. 4473–4482, 2016.

[54] C. S. A. G. M. S. V. Verdu M. O. Garcia and F. J. G. Franco, “Classification, filtering, and identification of electrical customer load patterns through the use of self-organizing maps,” *IEEE Transactions on Power Systems*, vol. 21, pp. 1672–1682, 2006.

[55] G. Chicco and I. S. Ilie, “Support vector clustering of electrical load pattern data,” *IEEE Transactions on Power Systems*, vol. 24, pp. 1619–1628, 2009.

[56] J. Y. L. M. Piao H. S. Shon and K. H. Ryu, “Subspace projection method based clustering analysis in load profiling,” *IEEE Transactions on Power Systems*, vol. 29, pp. 2628–2635, 2014.

[57] A. J. R. Reis and A. P. A. da Silva, “Feature extraction via multiresolution analysis for short-term load forecasting,” *IEEE Transactions on Power Systems*, vol. 20, pp. 189–198, 2005.

[58] J. Q. S. Z. M. L. H. Yang J. Zhang and Z. Y. Dong, “A practical pricing approach to smart grid demand response based on load classification,” *IEEE Transactions on Smart Grid*, vol. 9, pp. 179–190, 2018.

[59] B. T. H. H. Z. Ni S. Fu and X. Huang, “Experimental studies on indoor sign recognition and classification,” *2014 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)*, pp. 489–494, 2014.

[60] M. R. Haq and Z. Ni, “A composite extended nearest neighbor model for day-ahead load forecasting day-ahead load forecasting,” *2019 International Joint Conference on Neural Network (IJCNN’19)*, pp. 1–8, 2019.

[61] A. A. J. Buitrago and S. Asfour, “Electric load pattern classification using parameter estimation, clustering and aritificial neural networks,” *Internation Journal of Power and Energy Systems*, vol. 35, no. 4, pp. 167–174, 2015.

[62] T. Mitchell, “Electric load pattern classification using parameter estimation, clustering and aritificial neural networks,” *Machine Learning*, 1997.

[63] A. Majd and Aida, “K-nn based fault detection and classification methods for power transmission systems,” *Protection and Control of Modern Power Systems*, vol. 2, no. 1, pp. 32–42, 2017.

[64] C. H. K. Chang and C. Lin, "Coordinate descent method for large-scale l2-loss linear support vector machines," *Journal of Machine Learning Research*, vol. 9, no. 3, pp. 1369–1398, 2008.

[65] S. C. V. V and S. C. Srivastava, "Classification and modelling of loads in power systems using svm and optimization approach," *2015 IEEE Power Energy Society General Meeting*, pp. 1–5, 2015.

[66] "Minneapolis energy market data," 2018-2019.

[67] J. F. J. Kwac and R. Rajagopal, "Household energy consumption segmentation using hourly data," *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 420–430, 2014.

[68] J. H. Z. P. Z. Y. Xu Z. Y. Dong and K. P. Wong, "A reliable intelligent system for real-time dynamic security assessment of power systems," *IEEE Transactions on Power Systems*, vol. 27, pp. 1253–1263, 2012.

[69] "Australian energy market operator," 2012-2013.

[70] "Australian meteorological data," 2012-2013.

[71] "Open ei data: Brookings, sd, usa," 2017-2018.

[72] L. S. e. a. Huang. NE Shen. Z., "The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings: mathematical. Phy Eng Sci*, vol. 454, pp. 903–95, 1998.

[73] G. C. Xianzhao. Yang and H. Liu, "Improved empirical mode decomposition algorithm of processing complex signal for iot application," *International Journal of Distributed Sensor Networks*, vol. 18, no. 10, 2015.

[74] S. O. G. Hinton and Y. W. Teh, "A fast learning algorithm for deep belief networks," *Neural Computing*, vol. 18, no. 7, pp. 1527 –1554, 2015.

[75] G. Hinton and R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313(5786), 2015.

[76] E. Y. Y. Y. T. Yamashita M. Tanaka and H. Fujiyoshi, "To be bernouli or to be gaussian for a restricted boltzman machine," *22nd International Conference on Pattern RecognitionScience*, pp. 1520 –1525, 2014.

[77] G. Hinton, "Training products of experts by minimizing contarstive divergence," *Neural Computing*, vol. 14(8), pp. 1771 –1800, 2002.

[78] G. Hinton, "A practical guide to training restricted boltzman machine," *Neural Networks: Tricks of the Trade: Second Edition, Springer, Berlin, Heidelberg*, pp. 599 –619, 2012.

[79] M. R. I. Koprinska and V. Agelidis, "Correlation and instance based feature selection for electricity load forecasting," *Knowledge Based System*, vol. 82, pp. 29 –40, 2015.

[80] H. L. Z. S. Tinghui. Ouyang Yusen. He. and S. Baek, “A deep learning framework for short-term power load forecasting,” *CoRR*, vol. abs/1711.11519, 2017.